Psychometrika

CONTENTS

VOLUME FOURTEEN SEPTEMBER 1949 NUMBER THREE

PSYCHOMETRIKA—VOL. 14, NO. 3
SEPTEMBER, 1949

A NEW COEFFICIENT: APPLICATION TO BISERIAL CORRELATION AND TO ESTIMATION OF SELECTIVE EFFICIENCY*

HUBERT E. BROGDEN
PERSONNEL RESEARCH SECTION, A.G.O.

A coefficient of selective efficiency is proposed which can be usefully applied to selection problems involving the evaluation of the validity of (1) dichotomous predictors and (2) continuous predictors at a particular or at successive points of cut. Previously the author has shown that the product-moment correlation can be interpreted as a direct index of selective efficiency if the distribution forms of the criterion and the predictor are similar and the regression of the criterion on the predictor is linear. The coefficient proposed in the present article may be employed to evaluate selective efficiency of a continuous predictor at particular points of cut even when these assumptions are not tenable or are not applicable. It also is demonstrated that the proposed coefficient of selective efficiency may—with somewhat simpler and more generally applicable assumptions than those required in deriving the conventional formula—be employed as a substitute for the biserial correlation coefficient.

In a previous paper (1), the author demonstrated that the product-moment correlation coefficient could be interpreted as a direct index of selective efficiency. Selective efficiency was defined in that paper as the gain over random selection in mean criterion score achieved by selecting with the given predictor, divided by the gain that would have been achieved with perfect selection or selection on the criterion itself. This definition resulted from a logical examination of the objectives of selection and was regarded merely as one step in the solution to the problem of determining the proper interpretation of a product-moment validity coefficient. It was, however, briefly indicated in a footnote to this paper that this definition provided the basis for deriving a new biserial correlation coefficient.

Subsequently, it has seemed evident to the author that this definition also provided a direct basis for a coefficient with quite general application in estimating selective efficiency. It may, in other words, be applied in situations in which the assumptions required to demon-

*The opinions expressed in this paper are those of the author and are not to be interpreted as representing official Department of the Army policy.

strate that the product-moment correlation coefficient is an index of selective efficiency do not apply.

In the present paper, this coefficient of selective efficiency—which will be designated S—will be derived as a substitute for the biserial correlation coefficient. The argument for direct application of S as an index of selective efficiency will also be presented, together with some indication of situations in which it will prove useful.

A distinction between what is desired in a biserial correlation coefficient and what is desired in a coefficient of selective efficiency should be made explicit before these two separate problems are approached. A biserial correlation coefficient is an estimate of a product-moment coefficient, an indication of the product-moment coefficient that would have been obtained if the dichotomous variable were continuous. The significance and interpretation of the coefficient vary according to the goodness with which it estimates the product-moment and as the significance and interpretation of the product-moment coefficient vary. The problems involved in deciding whether to use and how to interpret such a coefficient are largely mathematical in nature and have to do with assumptions involved in the derivation of the biserial and in the derivation of the product-moment coefficient.

On the other hand, S as a coefficient of selective efficiency has been developed with the specific objective of measuring the efficiency with which a predictor will accomplish the objectives of selection. The absence of restrictive assumptions in its derivation allows general application within the general area of validation studies where a coefficient of selective efficiency is usually required.

In general, it is probably true that statistical formulas are not developed with the primary objective of providing interpretations most meaningful for a research worker having problems peculiar to a given area of research. The formula is more apt to be developed as an expression of certain mathematical relationships. In the derivation, assumptions—often highly limiting in nature—are introduced as necessary to the development of the given formula. Applications are sought at a later date. Very often it is found that the assumptions are so restrictive that the coefficient can legitemately be used in only a small proportion of certain types of applications. In other instances the coefficient may have legitimate application but may not provide the interpretation needed.

Derivation of the Conventional Biserial

It should help in further discussion if, first of all, we consider a derivation of the conventional biserial formula.

The following notation will be employed in this derivation and throughout this paper.

X — the dichotomous variable.
X' — a hypothetical continuous variable corresponding to X.
Y — the continuous variable in raw score form.
$Z_{X'}$, Z_Y — the standard score form (mean of zero, and $S.D.$ of one) of X' and Y, respectively.
P_X — proportion of cases in the upper category of X.
u — a subscript indicating that the symbol it modifies refers to those cases in the upper category of X.
v — a subscript indicating that the symbol it modifies refers to those cases in an upper category of Y equal in number to those in the upper category of X.

The three assumptions involved in this derivation of the conventional biserial formula are:

a. A continuous variable underlies the obtained dichotomous variable X.
b. The regression of the continuous variable, Y, on X' is linear. (We might note here that as a consequence of this assumption of linearity predicted Y values for any given X' value will fall on the regression line and, in addition, the predicted Y value for any linear combination of X' values will likewise fall on the regression line.)
c. The distribution of X' is normal.

Note: Nothing is directly assumed regarding the distribution of the continuous variable, Y, or regarding the regression of X' on Y.

The regression of Y on X' may be expressed as

$$\bar{Z}_Y = r_{X'Y} Z_{X'} . \tag{1}$$

The $\bar{Z}_Y$ of (1) is both the predicted value of Z_Y for the $Z_{X'}$ value of a given individual and the mean of the Z_Y values in an X' array. Equation (1) will hold in predicting an individual's Z_Y value from his $Z_{X'}$ value and any sum of different values since the regression of Y on X' is linear. Hence, with this assumption of linearity, Equation (1) may be employed to predict Y values for any individual above some point of cut on X', and the sum total for all individuals above the point of cut may be expressed as follows:

$$_u(\sum \bar{Z}_Y) = r_{X'Y}\, _u(\sum Z_{X'}). \tag{2}$$

The bar over Z in ${}_u(\Sigma\bar{Z}_Y)$ may be dropped, since in summing over a sub-population the errors of estimating individual Z_Y scores will "average out." Dividing both sides by N_u, the number above the point of cut, and reducing, we obtain

$$ {}_uM_{z_Y} = r_{X'Y}\,{}_uM_{z_{X'}} \tag{3} $$

or

$$ r_{X'Y} = {}_uM_{z_Y} / {}_uM_{z_{X'}} \,. \tag{4} $$

Thus, with the assumption indicated, $r_{X'Y}$ may be estimated as the ratio of two means. If X is dichotomous, ${}_uM_{z_{X'}}$ is unknown, but may be estimated if it is assumed that the continuous variable (X') is normally distributed. The mean of the tail of the normal curve is given by the formula Y/P_X, where Y is the height of the ordinate at the point of cut and P_X is the proportion of cases in the tail of the curve. After making the indicated substitution in (4) and converting to raw scores

$$ r_{\text{bis}} = \frac{{}_uM_Y - M_Y}{\sigma_Y} P_X/Y. \tag{5} $$

This is the conventional form of the biserial.

Limitations of the Conventional Coefficient; Derivation of Alternative Formulas

An elaboration of the implications of assuming normality of X' when Y is not normally distributed will aid in understanding why coefficients over unity are obtained with (5), even though the explicitly stated assumptions involved in its derivation are satisfied.

It is evident that the assumption of normality in the distribution of $Z_{X'}$ coupled with that of linear regression of Z_Y on $Z_{X'}$ requires that the $\bar{Z}_Y$ values predicted from this linear regression line be normally distributed. As a consequence any lack of normality in the distribution of Z_Y must be accounted for in the distribution of the errors of estimation $(Z_Y - \bar{Z}_Y)$. Since there are limits to the extent of the influence of the distribution of $(Z_Y - \bar{Z}_Y)$ on the distribution of Z_Y, particularly when $r_{X'Y}$ is high and the variance of $(Z_Y - \bar{Z}_Y)$ is small, there are apparently situations in which the two assumptions of linear regression of Z_Y on $Z_{X'}$ and normality of X' are mutually inconsistent. The appearance of biserial correlations above unity — which seem to occur with an anormal distribution of Z_Y—is undoubtedly re-

lated to inconsistency between these two assumptions. In the limiting case when the correlation is unity and when the regression of Z_Y on $Z_{X'}$, is linear, it is quite apparent that the distribution of $Z_{X'}$ and Z_Y must be of the same form. If the distribution of Z_Y is not normal it is equally apparent that at least in this limiting case the substitution of an estimate of $M_{Z_{X'}}$—such as Y/P—must bias r_{bis} as an estimate of r. Distortion or bias will very probably occur in other than the limiting case. While the presence of such bias is stressed, no attempt will be made here to trace the exact mechanism or to show the exact nature of its effect.

If it is agreed that the assumptions involved in the derivation of the conventional coefficient are unreasonable when the distribution of Y is not normal, examination of possible alternative assumptions should be of interest.

Two possibilities are suggested, both of which amount, in effect, to equating the distribution form of X' and Y. First of all it could be argued that if normality of the X' distribution is assumed, together with linear regression of Y on X', Y should also be assumed to be normally distributed. In effect, this is assuming bias in the units of the obtained Y distributions. The implications of such an assumption on computation are straightforward. The Y distribution may be normalized by use of appropriate tables. Additional computational labor required to normalize Y would not be excessive, especially if a considerable number of biserials were to be computed against each continuous variable. The actual computational process would involve determining the normalized values for the midpoints of the frequency intervals by the formula $(Y_1 - Y_2)/P_X$, where Y_1 and Y_2 are heights of the ordinates at the limits of the class interval as determined from the proportions exceeding these points. In the case of the intervals at the two extremes of the test, the above formula would reduce to Y/P_X. In both instances the P value in the denominator is the proportion of cases in that class interval. Since, for a distribution of such normalized values, M_Y of (5) would become M_{Z_Y} or zero and σ_Y would become σ_{Z_Y} and equal unity, (5) with Y in standard score form would become

$$r_{bis} = {}_uM_{Z_Y} \cdot P/Y. \tag{6}$$

While M_{Z_Y} is assumed to be zero and σ_{Z_Y} to be unity, it would probably be advisable, as a check, to calculate the mean and σ of the normalized values. With a small number of categories σ_{Z_Y} may fall below unity. In this event the value obtained from (6) should be di-

vided by the obtained σ_{z_Y} value to avoid the overestimation of r_{bis} that would result from underestimating σ_{z_Y}.

Exact correspondence of the distribution form of X' and Y is a second and possibly the most plausible of the several possible assumptions regarding the nature of the distribution of X' when the distribution of Y is not normal. With such an assumption, symmetry of the regression lines and of the frequency surface is plausible when the continuous variable is anormally distributed. This assumption leads to the derivation of S as an r_{bis} formula.

The specific assumptions involved in deriving the coefficient S are as follows:

(a) The dichotomous variable may be regarded as continuous.
(b) The regression of Y on X' is linear.
(c) The distribution form of X' is the same as the distribution form of Y.

The derivation follows that for the conventional coefficient through (4). Equation (4) will be repeated here for the convenience of the reader. In (4)

$$r_{X'Y} = {}_uM_{z_Y} / {}_uM_{z_{X'}} . \tag{4}$$

${}_uM_{z_Y}$ is the mean of the tail of the frequency distribution of Y. If, at this point, we assume that the X' has the same distribution form as Y, it follows that an equal tail of the X' distribution would have the same mean. Thus ${}_uM_{X'} = {}_vM_Y$, and substituting in (4) we have

$$r_{X'Y} = {}_uM_{z_Y} / {}_vM_{z_Y} . \tag{7}$$

Substituting raw scores, reducing, and designating the resulting coefficient S, we obtain

$$S = \frac{{}_uM_Y - M_Y}{{}_vM_Y - M_Y} . \tag{8}$$

Computations for (8) may be made rather rapidly. ${}_uM_Y$ may be computed from a frequency distribution of Y. If large numbers of correlations against a single criterion are being calculated, a table of such values may be prepared for each possible percentage above the cutting point on Y. It may, of course, be necessary to interpolate to obtain ${}_vM_Y$ for the appropriate percentile value. Linear interpolation would seem sufficiently accurate for most purposes.

A computational example is given below.

TABLE 1
Frequency Distribution of Cases in the Upper Category of the Dichotomous Variable and of the Total Group

Y	Upper Category	Total Group
7	2	3
6	2	5
5	3	10
4	3	15
3	2	10
2		5
1		3
N	12	51

COMPUTATIONS

1. $M_Y = 4$

2. ${}_uM_Y = [(2)\ 7 + (2)\ 6 + (3)\ 5 + (3)\ 4 + (2)\ 3]/12$
$= 4.91$ or the mean Y value of those cases in the upper category of the dichotomous variable.

3. ${}_vM_Y = [(3)\ 7 + (5)\ 6 + (4)\ 5.3^*]/12$
$= 6.02$ or the mean Y value of the 12 cases selected as highest on the continuous variable Y, 12 being the number in the upper category of the dichotomous variable.

4. $$S = \frac{4.91 - 4.00}{6.02 - 4.00} = .450$$

*To obtain ${}_vM_Y$ we need the average of the 12 cases highest on Y. This obviously includes the 3 cases having Y values of 7, the 5 cases having Y values of 6, and 4 of the 10 cases having Y values of 5. If we assume Y to be continuous, and the 10 cases having Y values of 5 to vary uniformly between 4.500 and 5.499, then the 4 cases highest within this interval can be presumed to range between 5.100, and 5.499, with a mid-point of 5.3.

S *as a Direct Index of Predictive Efficiency; Applications to Curvilinear Relations and to Distributions of Predictors and Criteria Which Are Not Normal*

In the introduction, it was stressed that S has direct meaning as a coefficient of selective efficiency apart from any relation which it may bear to the product-moment coefficient of correlation. S is justified directly from the definition of selective efficiency. It will be desirable consequently to review the nature of the selection process to determine exactly its objectives and to insure that the definition chosen is the most logical and appropriate.

By selection we refer to the process of identifying by means of predictor variables that portion of a general population which will be found to have high criterion scores. A predictor variable is usually employed in place of the criterion itself as a selector for reasons of economy or time or simply because the criterion values are not obtainable at the time at which selection must be made.

The nature of an adequate criterion is determined by the objectives of the organization for which selection is to be made. Thus, in an employment situation the objective would usually be increased quantity and quality of production; in a school or university, increased academic achievement. Criteria for the employment situations should, then, measure the differential effect of individual employees on the over-all productivity of the organization. In the school situation, academic achievement should be measured. Assuming that perfect criteria could be devised, the objectives of selection would be maximized by selection on the criterion itself.

Whatever the means of selection, the average gain achieved by selection is the difference between the mean criterion score of those selected and the mean criterion score of the total group from which they were selected. The latter gives the average or expected on-the-job productivity if selection were made at random from the total population. The former is the average on-the-job productivity of members of the selected group. This means, in other words, that with a given predictor or predictor battery employed at a given selection ratio the above difference in mean values shows the estimated absolute gain in productivity, per selected individual, resulting from the selection process. This value has meaning in its own right. However, it is a function of the selection ratio as well as the validity of the selection instrument. To obtain an index of selective efficiency of a predictor, the increase in productivity obtained with the predictor should be divided by the increase over random selection that would have been obtained with perfect (criterion) selection of the same number of applicants. This would give the *percentage* of possible gain actually achieved. If we translate this verbal definition of selective efficiency into statistical symbols, designating Y as the criterion, we obtain the coefficient

$$S = \frac{{}_uM_Y - M_Y}{{}_vM_Y - M_Y}.$$

${}_uM_Y$ would now be defined as the mean criterion score of those selected by the predictor and ${}_vM_Y$ as the mean criterion score of a group of the same number selected by the criterion.

It has already been shown how S may be directly computed, given test and criterion scores and the number or proportion of applicants to be hired. In addition it has been shown that the product-moment r is equal to S (1) if the regression of the criterion on the predictor is linear and the two distribution forms are the same. In stating that r equals S, we mean that this equality will hold no matter what point

of cut on the predictor is chosen for computation of S. There should be no problem, then, as to the evaluation of selective efficiency when these two assumptions are satisfied; the product-moment r is directly applicable.

While the assumption of linear regression of the criterion on the predictor is readily understood, the assumption of equal distribution forms implies acceptance of certain principles which should be clarified. Such clarification is important because the applicability of either r or S as an index of selective efficiency hinges upon acceptance of these principles. First of all it is implied that the criterion distribution has meaning in its own right or that the criterion scale units represent equal increments of the variable measured. Where the criterion scale consists of such production units as number of objects produced or number of errors made, this assumption is apparently quite legitimate. Errors or objects produced are units having definite and standard significance relative to the objective of the selection program—improvement of the efficiency of operation of the organization for which selection is made. If ratings are employed as criteria, the experimenter will have to decide from knowledge of the particular scale whether or not sufficient bias in scale units exist as to make this assumption unjustifiable. Unfortunately, it will probably be impossible to arrive at a definite decision.

To digress for a moment, we might note that a coefficient dependent upon the assumption of equal scale units has definite advantages over coefficients such as the conventional biserial which tend toward biased estimates of validity without normality of the criterion distribution. (See discussion on page 172). From the viewpoint of the objective of selection, the need for normalizing anormal criterion distributions before an index of selective efficiency will properly apply is equivalent to the necessity of converting to non-meaningful units before selective efficiency can be determined. The experimenter is faced with the dilemna of being unable to determine selective efficiency or of applying a coefficient which will result in a distortion of the proper answer to his problem.

A second implication of the assumption of equal distribution forms is that the predictor scale units have no direct meaning for the purpose of evaluating selective efficiency. Thus, if the distribution form of the predictor is the same as that of the criterion—and the regression of the criterion on the predictor is linear—the product-moment coefficient may be appropriately employed as an index of selective efficiency. When r is not appropriate, S should be employed to determine selective efficiency at various cutting points throughout the range of predictor scores. From the formula for S it can be seen

that the distribution form of the predictor is, in the latter instance, ignored.

This point needs little elaboration. We might, however, note that the predictor distribution form necessary to use of r as an index of selective efficiency has no necessary relation to the distribution form which will provide maximum predictive efficiency. The latter problem, in the case of test scores which are sums of dichotomous items, is a problem in the proper distribution of item difficulties.

When the assumptions of linear regression and equal distribution forms are known not to be true, or suspected not to be true, S will provide directly the desired index of selective efficiency for particular points of cut or particular selection ratios. S may also be employed in certain additional situations where the product-moment correlation is obviously not applicable. Thus, with dichotomous predictors, S provides, without the assumptions involved in its derivation as a biserial correlation coefficient, a direct index of selective efficiency. This is directly evident only if the proportion in the upper category of the predictor corresponds to the proportion selected. However, S may be readily adapted to estimating selective efficiency when these two proportions do not correspond. This may be done by redefining the number of cases selected on the criterion, in computing ${}_{v}M_{Y}$, as the number of applicants it is desired to select rather than the number in the upper category of the dichotomous variable.

If multiple cutting scores have been set for several predictors, S may be employed as an index of the selective efficiency obtained when a battery of tests is utilized in this manner.*

Application of S with dichotomous predictors or in the case of multiple cutting scores is not in need of further elaboration. Its application to continuous predictors when r is not applicable because of non-linear regression or inequality of distribution forms will bear further discussion.

When the product-moment correlation coefficient is not applicable, its inapplicability means in effect that r would not equal S if the latter were computed for all points of cut on the predictor. It follows consequently that these assumptions may be tested by computing S for points of cut covering the range of the predictor. Such computations would not only test the applicability of r as an index of selective efficiency but would indicate the extent of error introduced by employing r as an index of selective efficiency and allow estimation of improvement in selective efficiency resulting from choice of

*For such application ${}_{u}M_{Y}$ is the mean criterion score of those "accepted" after application of the multiple cutting score procedure, while ${}_{v}M_{Y}$ is the mean criterion score of a group of comparable size selected on the criterion itself.

particular cutting points. If several alternative predictors, or predictor composites, are available, that one most suitable for selection at a predetermined selection ratio could be chosen. With a predetermined selection ratio, S may also prove of value in combining predictor variables into a composite. The exact manner of its application to this problem is not clear. It is apparent from perfunctory review of the derivation of partial regression coefficients that it cannot be readily proved that S may be employed for computing validity coefficients to be used in multiple regression analysis. However, if several weighted combinations of predictors were tried, S could be employed to determine that yielding the highest predictive efficiency for the given predetermined selection ratio.

A plot of particular S values against the percentage above the point of cut involved, which might be termed a curve of selective efficiency, is suggested as an aid in determining whether or not descrepancies between S and r show systematic trends or merely chance deviations. Such a curve should be useful in deciding upon the selection ratio and in choosing from several possible predictors that one most suited to selection at a predetermined selection ratio.

A curve of selective efficiency, as determined by computing S at various points of cut, does not provide information corresponding either to that provided by eta or to that provided by fitting a curvilinear regression line. Eta provides a single estimate of correlation for the entire range of the scatter-plot. A curvilinear regression line provides both such an estimate of correlation and the predicted criterion score for any given test score. In actual practice, however, it is usually not desired to select a group having a particular test score but a group above some given test score. Additional computation would be necessary to obtain the mean score of such a group from a curvilinear regression line. The coefficient S provides directly the information desired in evaluating selective efficiency and should, in the author's opinion, be preferable to the alternative mentioned in evaluating selective efficiency whenever curvilinear regression of the criterion on the predictor is known or suspected to exist.

A point previously made might be stressed again in this connection. Curvilinear regression may be and sometimes is linearized by alterations in the predictor scales. An equation may be employed for this purpose or each predictor value may be assigned its actual $\overline{Y}$ value. Such converted predictor scales will maximize the product-moment r. The S values computed at various points of cut will, however, be unaffected, since S is computed entirely from criterion scores. This further implies that solution to the problem of obtaining maximum selective efficiency of a composite or weighted sum for a pre-

determined selection ratio cannot be fully solved by altering the scale values of the component predictors.

Curvilinear regression lines have not been widely used in practice. This is undoubtedly due in part to the labor required in their application. It has often been found, however, that even with apparently marked curvilinear regression little improvement is found over the predictive efficiency obtained with the best straight-line approximation to the curvilinear regression line. The author would agree and even emphasize the poor expectancy of improved efficiency from application of curvilinear regression when it is desired to predict over the entire range of the variables in question. When it is desired to employ a definite selection ratio (or a definite point of cut) or when it is possible to modify the selection ratio to take advantage of any variation in selective efficiency that may be discovered, more fruitful results may be expected. A correlation coefficient computed from eta or from a curvilinear regression line, since it does provide an "average" selective efficiency coefficient, conceals and ignores these differences in selective efficiency which may be identified by use of S and used to advantage.

The point made above may be effectively illustrated by a numerical example. In Figure 1 we have a scatter plot with the regression of Y on X indicated, with r and eta computed for the entire range and S determined for various points of cut. It will be noted that while the regression appears to be definitely curvilinear, r and eta do not differ markedly. However, the differences in the value of S for the several points of cut are of considerable magnitude.

To utilize with confidence the differences between the values of S at different points of cut, the number of cases would have to be much greater than in this example. The technique of cross-validation should probably be applied in such circumstances in order to obtain an unbiased estimate of validity if selection of the particular point of cut were dependent upon the obtained values of S. If the point of cut were predetermined, cross-validation would be unnecessary.

A special problem in which S has particular application occurs in estimating selective efficiency of tests constructed for the particular purpose of selecting at a predetermined point of cut. In obtaining an efficient test for that purpose items should be selected which have P values approximately equal to the per cent of the population to be eliminated (2). Such a selection will obviously lead to a distribution of test scores whose mean, standard deviation, and distribution form have no relation to the distribution of "true" ability in the function measured. Since the product-moment r will be influenced by the standard

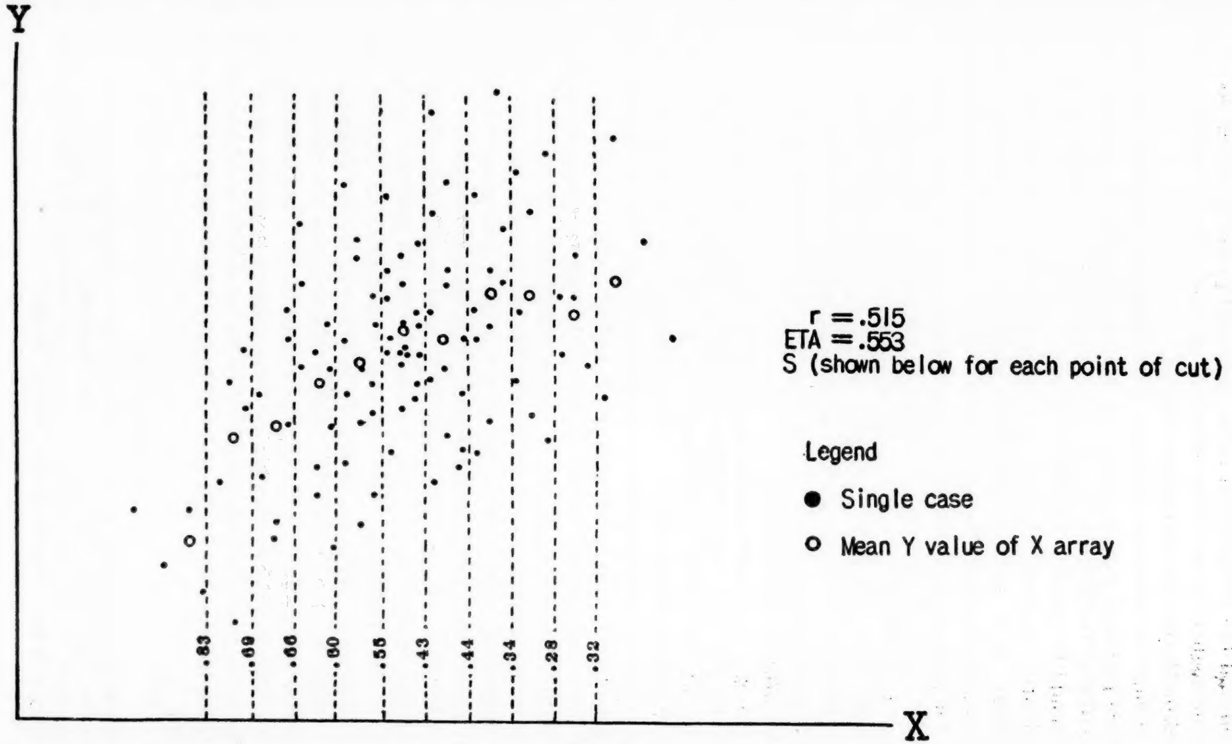


FIGURE 1

A Hypothetical Scatter Plot, Showing Mean Y Values for X Arrays, with Computed Values of r, Eta and S (For the Various Points of Cut on X Indicated by the Dotted Lines).

deviation and the distribution form, it should probably never be employed for evaluating selective efficiency of such a test. S will in such circumstances give the evaluation desired.

If a curve of selective efficiency were obtained by computing S at successive points of cut, it would not only be possible to measure the validity of S at the point or in the area for which the test is designed, but it should be possible to detect "wasted" efficiency in the sense of discrimination in areas where such a specialized test was not intended to function. It is realized that there would be, at the present time, little basis for deciding the optimal form of the curve of selective efficiency for such a test. Probably a test with items of low reliability would show a shallow curve of selective efficiency, while a test with items of high reliability would be more markedly curvilinear and would show a more definite optimal point. Any decision as to deletion of items made on the basis of such a curve of selective efficiency should probably be checked empirically by recomputing the curve of selective efficiency after item selection.

From the two assumptions required to show equality between r and S at all points of cut, it is apparent that variation of S for different points of cut may be due to differences in the distribution forms of the predictor and criterion as well as to curvilinear regression of Y on X. Of course, either or both of these two observed phenomena may be due in turn to other characteristics of the correlation surface.

If the criterion distribution is highly skewed or otherwise lacking in normality, the possibility of obtaining differential selective efficiency at different points of cut is of considerable interest. As in the case of the suggested application of S in curvilinear regression, the extent of the differences may be calculated for each possible selection instrument or battery and used to advantage in selecting tests or determining selection ratios. In practice, as was mentioned before, bias in the criterion scale units may mislead the investigator in this respect. Where production units are available as criteria, the investigator can often accept the distribution as having meaning for his purpose—regardless of the relation of production units to any hypothetical underlying ability. When ratings or achievement test scores are the criteria to be predicted, the obtained curve of selective efficiency will have to be interpreted in the light of known biases in the scale units involved.

REFERENCES

1. Brogden, Hubert E. On the interpretation of the correlation coefficient as a measure of predictive efficiency. *J. educ. Psychol.*, 1946, 37, 65-76.
2. Richardson, M. W. The relation between the difficulty and the differential validity of a test. *Psychometrika*, 1936, 1, 33-49.

PSYCHOMETRIKA—VOL. 14, NO. 3
SEPTEMBER, 1949

FURTHER NOTES ON DIFFERENCES BETWEEN PERCENTAGES

FRANCES SWINEFORD
EDUCATIONAL TESTING SERVICE

In certain problems dealing with percentage differences, it frequently happens that one is interested in a critical value different from zero rather than in the usual hypothesis of no population difference. Moreover, one may be concerned with differences greater than (or less than) a given magnitude, so that only one tail of the distribution of chance values is used. Tables are here presented to facilitate evaluating results of such problems. One per cent and 5 per cent values can be read from the tables, such values corresponding to the more usual 2 per cent and 10 per cent confidence limits.

An earlier article* presented graphs and a table to assist in the determination of the statistical significance of the difference between two percentages or the determination of the sample size necessary to establish significance of a difference of given magnitude.

The writer has recently been concerned with another problem, related to the foregoing one, but dealing with different null hypotheses.† It frequently happens that one is not so much interested in whether a difference is significantly greater than zero as in whether it might represent a chance variation from some hypothetical value other than zero. Thus, for example, in a problem of item analysis suppose a critical value of .30 to have been set for an acceptable difference between proportions of criterion groups answering an item correctly. If there were 100 cases in each group, an obtained difference of .25 would be considered statistically significant, but it could also be regarded as a chance deviation from .30 until further evidence concerning it should be secured. When such an item appears on other grounds to be a desirable item, it need not be discarded until it is shown to yield a difference significantly under .30.

In general, it it convenient to consider such problems in terms of sample size. In the foregoing instance the problem is one of determining the sample size which would be necessary to reject the null

*Swineford, Frances. Graphical and tabular aids for determining sample size when planning experiments which involve comparisons of percentages. *Psychometrika*, 1946, 11, 43-49.

†This problem was suggested by Dr. L. R Tucker of the Educational Testing Service.

hypothesis that the population difference is .30 (or greater). Since this amounts to a consideration of only one tail of the normal distribution, the 5 per cent and 1 per cent limits correspond to 10 per cent and 2 per cent confidence limits in their usual sense.

Two tables have been prepared to simplify the task of computing the critical ratios or their equivalent. It is assumed that the samples to be compared are of equal size, N. The standard errors of the two proportions are

$$\sigma_{p_1} = \sqrt{\frac{p_1 q_1}{N}} \quad \text{and} \quad \sigma_{p_2} = \sqrt{\frac{p_2 q_2}{N}}, \tag{1}$$

from which the standard error of the difference, $d = p_1 - p_2$, is

$$\sigma_d = \sqrt{\frac{p_1 q_1 + p_2 q_2}{N}}. \tag{2}$$

A first approximation to σ_d is obtained by substituting for p_1 and p_2 their average, p:

$$\sigma_d = \sqrt{\frac{2pq}{N}} \tag{3}$$

Five per cent of the area under the normal curve lies above 1.64485σ, and 1 per cent, above 2.32635σ. Thus for an observed difference, d_o, and a theoretical difference, d_t, we may write

$$\text{5 per cent:} \quad d_o - d_t = 1.64485 \sqrt{\frac{2pq}{N}} \ ; \tag{4}$$

$$\text{1 per cent:} \quad d_o - d_t = 2.32635 \sqrt{\frac{2pq}{N}}. \tag{5}$$

Solving for N in (4) and (5), respectively, gives

$$N = \frac{5.4111pq}{(d_o - d_t)^2} \text{ for the 5 per cent case,} \tag{6}$$

and

$$N = \frac{10.8238pq}{(d_o - d_t)^2} \text{ for the 1 per cent case.} \tag{7}$$

Since the ratio of the 1 per cent expression to the 5 per cent expression is 2.0003, only one set of values of N will be recorded in Table 1—those for the 1 per cent case. Halving any value will then give the 5 per cent figure.

Linear interpolation can be used in all parts of the table. In the lowest section, however, a small error is introduced thereby, but in no instance will the error exceed six cases in any row, nor will it exceed two cases in any column. Horizontal interpolation underestimates N, and vertical interpolation overestimates N.

The second table provides corrections for the simplification introduced in formula (3). This formula is appropriate only in connection with the null hypothesis of no difference in the population. The tabled values are suggested for use with the null hypothesis that the population difference is some value other than zero. The discrepancy between values from formulas (2) and (3) increases slightly as p moves away from .50, and it increases substantially with increasing values of the hypothetical difference, d_t. The entries in Table 2 are the ratios, $(p_1q_1 + p_2q_2)/pq$, by which the N's from Table 1 should be multiplied.

Two examples will serve to illustrate the use of the tables. Suppose the percentages for two equal groups to be 76 and 72 with a p of .74. Suppose one wishes to know whether one may regard as tenable the hypothesis that the population difference is 15 per cent. Entering Table 1 with $p = .74$ and $|d_o - d_t| = |.04 - .15| = .11$, we obtain an N of 172. Entering Table 2 with $p = .74$ and $d_t = .15$, we obtain a factor of .971. The product of these two figures, $.971 \times 172$, is 167, and we conclude that if the number of cases in each group is no greater than 167 any hypothesis up to a difference of 15 per cent is acceptable at the 1 per cent level of confidence. If the number of cases in each group exceeds 167, the null hypothesis ($d_t = .15$) is rejected at the 1 per cent level. Half this number of cases or 84 cases are sufficient to draw the same conclusion at the 5 per cent level of confidence.

For a second example, suppose the two proportions to be .43 and .27 with a p of .35 and a difference of .16. Assume that we wish to know how large each group must be in order that we may reject at the 1 per cent level of confidence any hypothetical difference of .10 or less. This, of course, is another way of saying that if the true difference is .10 the probability that the obtained difference will exceed .16 is only .01. In this instance Table 1 is entered with $p = .35$ and $|d_o - d_t| = |.16 - .10| = .06$, which gives an N of 684. In Table 2 is found the factor .989, and $.989 \times 684 = 676$. For $N \geq 676$, therefore, we may reject at the 1 per cent level hypothetical differences of .10 or under.

TABLE 1
1% Values of N for Selected Values of $d_o - d_t$ and p

$\lvert d_o - d_t \rvert$	p								
	.50	.55 .45	.60 .40	.65 .35	.70 .30	.75 .25	.80 .20	.85 .15	.90 .10
.135	148	147	143	135	125	111	95	76	
.130	160	159	154	146	134	120	102	82	
.125	173	171	166	158	145	130	111	88	
.120	188	186	180	171	158	141	120	96	
.115	205	203	196	186	172	153	131	104	
.110	224	221	215	203	188	168	143	114	
.105	245	243	236	223	206	184	157	125	
.100	271	268	260	246	227	203	173	138	97
.095	300	297	288	273	252	225	192	153	108
.090	334	331	321	304	281	251	214	170	120
.085	375	371	360	341	315	281	240	191	135
.080	423	419	406	385	355	317	271	216	152
.078	445	440	427	405	374	334	285	227	160
.076	468	464	450	426	394	351	300	239	169
.074	494	489	474	450	415	371	316	252	178
.072	522	517	501	475	438	391	334	266	188
.070	552	547	530	503	464	414	353	282	199
.068	585	579	562	533	492	439	375	298	211
.066	621	615	596	565	522	466	398	317	224
.064	661	654	634	601	555	495	423	337	238
.062	704	697	676	641	591	528	451	359	253
.060	752	744	722	684	631	564	481	383	271
.058	804	796	772	732	676	603	515	410	290
.056	863	854	828	785	725	647	552	440	311
.054	928	919	891	844	779	696	594	473	334
.052	1001	991	961	911	841	751	640	510	360
.050	1082	1072	1039	985	909	812	693	552	390

TABLE 2

Values of $\frac{p_1q_1 + p_2q_2}{pq}$ for Selected Values of d_t and p

d_t	p								
		.55	.60	.65	.70	.75	.80	.85	90
	.50	.45	.40	.35	.30	.25	.20	.15	.10
.10	.990	.990	.990	.989	.988	.987	.984	.980	.972
.15	.978	.977	.977	.975	.973	.970	.965	.956	
.20	.960	.960	.958	.956	.952	.947	.938	.922	
.25	.938	.937	.935	.931	.926	.917	.902		
.30	.910	.909	.906	.901	.893	.880	.859		
.35	.878	.876	.872	.865	.854	.837			
.40	.840	.838	.833	.824	.810	.787			
.45	.798	.795	.789	.777	.759				
.50	.750	.747	.740	.725	.702				

PSYCHOMETRIKA—VOL. 14, NO. 3
SEPTEMBER, 1949

VARIATION OF THE STANDARD ERROR OF MEASUREMENT

WILLIAM G. MOLLENKOPF*
PRINCETON UNIVERSITY AND
EDUCATIONAL TESTING SERVICE

As usually interpreted, the standard error of measurement is assumed to be constant throughout the test-score range. In this investigation the standard error of measurement was assumed to be not higher than a second-degree function of the test score. By conceiving a test score to be made up of the scores on two parallel tests, an equation was derived for predicting the standard error of measurement from the test score. In the derivation the corresponding first four moments of the score distributions for the parallel tests were assumed to be identical, and certain errors of estimate involved in predicting the second test score from the first were assumed to be uncorrelated with powers of the score on the first test. An empirical verification was carried out, using nine synthetic tests and a 1000-case sample, and showed good agreement between predicted and observed results. The findings indicated that the standard error of measurement was constant only for a symmetrical, mesokurtic distribution of scores.

Introduction

The standard error of measurement is a basic concept in the theory of mental tests. As typically defined by the relationship

$$\sigma_{\text{meas}_x} = \sigma_x \sqrt{1 - R_{xx}}, \tag{1}$$

where

R_{xx} = the reliability and
σ_x = the standard deviation of the test,

it is an average or over-all measure indicating the extent to which actual scores made by a group of persons would vary about their true scores if these persons were to be given a large number of parallel tests. When N individuals have been given K parallel tests, the standard error of measurement might be found by taking for each

*This study was carried out while the author was a National Research Council Predoctoral Fellow in psychology at Princeton University.

The author wishes to express his appreciation for the guidance given by his thesis adviser, Professor Harold Gulliksen. He wishes also to acknowledge his gratitude to the Educational Testing Service for extensive assistance in the empirical phase of the study, and to Dr. Ledyard Tucker for suggesting efficient methods of handling special computational problems.

individual the difference between his observed score on each test and his true score, squaring each of these differences, summing over tests and individuals, dividing the result by the product of the number of tests and the number of individuals, and then taking the square root of the quotient. In symbols,

$$\sigma_{\text{meas}_x} = \sqrt{\frac{\sum_{i=1}^{N} \sum_{j=1}^{K} (x_{ij} - t_i)^2}{KN}}, \tag{2}$$

where

x_{ij} = the observed score of individual i on test j,
t_i = the true score for individual i,
N = the number of individuals, and
K = the number of parallel tests.

The statement of the standard error of measurement given in (2) can be shown to be equivalent to that given in (1) when a true score for an individual is defined as the mean of his observed scores on a large number of parallel tests, that is, when

$$t_i = \frac{\sum_{j=1}^{K} x_{ij}}{K}. \tag{3}$$

In making use of the standard error of measurement in interpreting test scores, the assumption is made that the variability about his true score of a person's observed scores on many parallel forms would be the same, regardless of whether this particular person's score happened to be low, average, or high. A pertinent question can therefore be raised as to the constancy or non-constancy of the standard error of measurement over the test-score range. The present investigation was concerned with the manner in which the standard error of measurement varied with magnitude of test score, for each of several types of score distributions. The problem was first treated analytically by mathematical means; an equation for predicting the size of the standard error of measurement from the test reliability and parameters of the test-score distribution was developed; and finally an empirical verification was carried through for several types of test-score distributions.

Theoretical Analysis of the Problem

Consider the problem of finding the standard error of measure-

ment for a test score which is the sum of the scores on two parallel tests. Then this total test score, x_i, is defined by the equation

$$x_i = x_{i1} + x_{i2}, \tag{4}$$

where x_{i1} and x_{i2} = the deviation scores of individual i on the parallel tests 1 and 2.

Using the Spearman-Brown formula we may express R_{xx}, the reliability of the total test, in terms of the correlation between the two parallel halves, r_{12}, as follows:

$$R_{xx} = \frac{2\, r_{12}}{1 + r_{12}}. \tag{5}$$

The standard deviation of the total test is related to that of each half by the equation

$$\sigma_x = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2 + 2\sigma_{x_1}\sigma_{x_2} r_{12}}, \tag{6}$$

where σ_{x_1} and σ_{x_2} = the standard deviations of the two parallel tests.

For parallel tests $\sigma_{x_1} = \sigma_{x_2}$, so we may rewrite (6) as

$$\sigma_x = \sigma_{x_1}\sqrt{2(1 + r_{12})}. \tag{7}$$

Upon substituting from (5) and (7) in (1) and simplifying, we find

$$\sigma_{\text{meas}_x} = \sigma_{x_1}\sqrt{2(1 - r_{12})}. \tag{8}$$

The right-hand side of (8) can be recognized as $\sigma_{x_1 - x_2}$, since $M_{x_1} = M_{x_2}$ for parallel tests. We may therefore write

$$\sigma_{\text{meas}_x} = \sqrt{\frac{\sum_{i=1}^{N} (x_{i1} - x_{i2})^2}{N}}. \tag{9}$$

Consequently, to find the standard error of measurement for a test the score on which is the sum of the scores on two parallel tests, one may take the sum of the squares of the differences between corresponding individual scores, divide by N, and extract the square root.

The assumption usually made concerning the standard error of measurement is that it is a constant. In algebraic language, this can be expressed

$$y_i = k,$$

in which y_i represents the standard error of measurement and k is

a constant. In the present study a hypothesis of a more general relationship was set up. Specifically, it was assumed that an equation of not higher than the second degree was adequate for expressing the relationship between the total test score as the independent variable and the error of measurement as the dependent variable. In symbols, this is

$$\text{Error} = ax^2 + bx + c\,,$$

in which x represents a total test score and a, b, and c are parameters to be determined. It is to be noted that the assumption that the relationship is not higher than a second-degree one does *not* preclude the possibility that the actual relationship may be a linear one with any slope, including zero, for either a or b or both may turn out to be zero. (The usual assumption of a constant standard error corresponds to a zero-slope straight line of relationship between error and total test score.)

Preliminary Statements

In the following analytical development, let

$$y_i = (x_{i1} - x_{i2})^2,$$

where x_{i1} and x_{i2} are defined as for equation (4). The square of the difference between the two parallel-test scores represents the square of the standard error of measurement and is designated y_i. (To avoid radicals, it will be convenient to work with this square.)

In addition to the assumptions already stated ($M_{x_1} = M_{x_2}$ and $\sigma_{x_1} = \sigma_{x_2}$) the following assumptions will also be made concerning scores on the two parallel tests:

$$\alpha_3' = \alpha_3''\,, \text{ where } \alpha_3' = \frac{\sum x_1^3 *}{N\,\sigma_{x_1}^3} \text{ and } \alpha_3'' = \frac{\sum x_2^3}{N\,\sigma_{x_2}^3}\,,$$

and

$$\beta_2' = \beta_2''\,, \text{ where } \beta_2' = \frac{\sum x_1^4}{N\,\sigma_{x_1}^4} \text{ and } \beta_2'' = \frac{\sum x_2^4}{N\,\sigma_{x_2}^4}\,.$$

It is evident that the foregoing assumptions are equivalent to assuming that the corresponding first four moments of the score distributions on the two parallel tests are identical.

*When the limits of a summation are not indicated, the summation is over i from 1 to N.

From the assumptions which have been made it follows that

$\sum x_1^3 = \sum x_2^3$, $\sum x_1^4 = \sum x_2^4$, $\sum x_1 = 0$, $\sum x_2 = 0$, and $\sum x = 0$.

Least Squares Solution

The next step is to determine the parameters in the equation

$$\dot{y}_i = ax_i^2 + bx_i + c\,, \tag{10}$$

where $\dot{y}_i$ represents the error as estimated from the test score. To determine the coefficients a, b, and c by the method of least squares, the quantity $\sum\,(y_i - \dot{y}_i)^2$ should be made a minimum.

Clearly

$$\sum\,(y_i - \dot{y}_i)^2 = \sum[y_i - (ax_i^2 + bx_i + c)]^2.$$

Taking in turn partial derivatives with respect to a, b, and c of the quantity to the right of the equals sign, setting each of the resulting expressions equal to zero, and carrying out slight simplification yields

$$\begin{aligned} &a\sum x^4 + b\sum x^3 + c\sum x^2 = \sum x^2 y\,,\\ &a\sum x^3 + b\sum x^2 + c\sum x = \sum xy\,, \quad \text{and}\\ &a\sum x^2 + b\sum x + cN = \sum y\,. \end{aligned}$$

The solution of these expressed in determinantal form is

$$a = \frac{\begin{vmatrix} \sum x^2y & \sum x^3 & \sum x^2 \\ \sum xy & \sum x^2 & \sum x \\ \sum y & \sum x & N \end{vmatrix}}{\begin{vmatrix} \sum x^4 & \sum x^3 & \sum x^2 \\ \sum x^3 & \sum x^2 & \sum x \\ \sum x^2 & \sum x & N \end{vmatrix}}, \quad b = \frac{\begin{vmatrix} \sum x^4 & \sum x^2y & \sum x^2 \\ \sum x^3 & \sum xy & \sum x \\ \sum x^2 & \sum y & N \end{vmatrix}}{\begin{vmatrix} \sum x^4 & \sum x^3 & \sum x^2 \\ \sum x^3 & \sum x^2 & \sum x \\ \sum x^2 & \sum x & N \end{vmatrix}}, \quad c = \frac{\begin{vmatrix} \sum x^4 & \sum x^3 & \sum x^2y \\ \sum x^3 & \sum x^2 & \sum xy \\ \sum x^2 & \sum x & \sum y \end{vmatrix}}{\begin{vmatrix} \sum x^4 & \sum x^3 & \sum x^2 \\ \sum x^3 & \sum x^2 & \sum x \\ \sum x^2 & \sum x & N \end{vmatrix}}.$$

Preliminary Derivations

Our purpose in this theoretical development is to express the parameters a, b, and c of equation (10) in terms of the moments of the total score distribution and the reliability of the test. Upon examination of the terms involved in the determinants stated above, it is evident that the quantities $\sum x^4$, $\sum x^3$, $\sum x^2$, and $\sum x$ can be readily restated in terms of moments of the total-score distribution. However, expressions must be derived for restating the quantities $\sum x^2 y$, $\sum xy$, and $\sum y$ in terms of the moments of the total-score distribution and the reliability of the test.

Before new expressions for $\sum x^2y$, $\sum xy$, and $\sum y$ can be secured, several relationships must be determined. These are: (1) the relation of the correlation between parallel forms to the reliability of the total test; (2) the relation of the variance of the half-test (parallel form) to the variance of the total test; (3) the relation of the skewness of the total form, α_3, to the skewness of the half-test, α_3'; and (4) the relation of the kurtosis of the total test, β_2, to the kurtosis of the half-test, β_2'.

Relation of Correlation between Parallel Forms to Reliability of Total Test: When equation (6) is solved for r_{12}, we find

$$r_{12} = \frac{R}{2-R}. \tag{11}$$

Relation of Variance of Half-test (Parallel Form) to Variance of Total Test: After squaring both sides of equation (7), we may write

$$\sigma_x^2 = 2\,\sigma_{x_1}^2\,(1 + r_{12}). \tag{12}$$

With r_{12} in (12) replaced by its value as given in (11), this becomes, after simplification,

$$\sigma_x^2 = \sigma_{x_1}^2 \left(\frac{4}{2-R}\right). \tag{13}$$

Or, solved for the variance of the half-test, this is

$$\sigma_{x_1}^2 = \sigma_x^2 \left(\frac{2-R}{4}\right). \tag{14}$$

Relation of α_3 to α_3': Cubing both sides of (4) and summing over i, we have

$$\sum x_i^3 = \sum x_{i1}^3 + 3\sum x_{i1}^2 x_{i2} + 3\sum x_{i1} x_{i2}^2 + \sum x_{i2}^3. \tag{15}$$

In Figure 1 the straight line of best fit (in the least squares sense) between x_1 and x_2 has been represented. Its equation is

$$x_2 = \frac{\sigma_{x_2}}{\sigma_{x_1}} r_{12}\, x_1. \tag{16}$$

Let

$$x_{2_{ig}} - r_{12}\frac{\sigma_{x_2}}{\sigma_{x_1}} x_{1_g} = x'_{2_{ig}}, \tag{17}$$

where the subscript g refers to a column ($g = 1, \cdots, G$) in the two-way scatter plot of Figure 1 and the subscript i refers to a person ($i = 1, \cdots, N_g$ for column g). In this notation $\sum x_{i1}^2 x_{i2}$ becomes

$$\sum_{g=1}^{G} x_{1_g}^2 \sum_{i=1}^{N_g} x_{2_{ig}} , \tag{18}$$

with x_1 and x_2 defined as for (4).

After solving (17) for $x_{2_{ig}}$ and then summing over i in column g, we have

$$\sum_{i=1}^{N_g} x_{2_{ig}} = \sum_{i=1}^{N_g} x'_{2_{ig}} + r_{12} \frac{\sigma_{x_2}}{\sigma_{x_1}} N_g \, x_{1_g} . \tag{19}$$

We may now use this equivalent in (18) and write

$$\sum_{i=1}^{N} x_{i1}^2 x_{i2} = \sum_{g=1}^{G} x_{1_g}^2 \sum_{i=1}^{N_g} x'_{2_{ig}} + r_{12} \frac{\sigma_{x_2}}{\sigma_{x_1}} \sum_{g=1}^{G} N_g \, x_{1_g}^3 . \tag{20}$$

We shall now assume that the mean of the errors of estimate for a column, $\overline{x'_{2_g}}$, is not correlated with $x_{1_g}^2$, that is, that

$$r_{x_{1_g}^2 , \overline{x'_{2_g}}} = 0 , \tag{21}$$

where

$$\overline{x'_{2_g}} \equiv \frac{\sum_{i=1}^{N_g} x'_{g}}{N_g} , \tag{22}$$

and

$$r_{x_{1_g}^2 , \overline{x'_{2_g}}} \equiv \frac{\sum_{g=1}^{G} N_g \, (x_{1_g}^2) \, (\overline{x'_{2_g}}) - N \, M_{x_{1_g}^2} \, M_{\overline{x'_{2_g}}}}{N \, \sigma_{x_{1_g}^2} \, \sigma_{\overline{x'_{2_g}}}} . \tag{23}$$

Under these conditions

$$\sum_{g=1}^{G} N_g (x_{1_g}^2)(\overline{x'_{2_g}}) = N\, M_{x_{1_g}^2}\, M_{\overline{x'_{2_g}}}$$

$$= N \left[\frac{\sum_{g=1}^{G} N_g\, x_{1_g}^2}{\sum_{g=1}^{G} N_g} \right] \left[\frac{\sum_{g=1}^{G} N_g\, \overline{x'_{2_g}}}{\sum_{g=1}^{G} N_g} \right] . \tag{24}$$

The quantity $\sum_{g=1}^{G} N_g\, \overline{x'_{2_g}}$ is equivalent to $\sum_{g=1}^{G} \sum_{i=1}^{N_g} x'_{2_g}$. Since this is the sum over the entire two-way table of the deviations from the line of best fit, it is zero. Hence with the assumption in (21) we may write

$$\sum_{g=1}^{G} N_g\, x_{1_g}^2\, \overline{x'_{2_g}} = 0. \tag{25}$$

From (22) and (25) we can now write

$$\sum_{g=1}^{G} x_{1_g}^2 \sum_{i=1}^{N_g} x'_{2_{ig}} = 0. \tag{26}$$

Noting that $\sum_{g=1}^{G} N_g\, x_{1_g}^3 = N\, \alpha'_3\, \sigma_{x_1}^3$, we may now rewrite (20) as

$$\sum_{i=1}^{N} x_{i1}^2\, x_{i2} = N\, r_{12}\, \sigma_{x_1}^2\, \sigma_{x_2}\, \alpha'_3. \tag{27}$$

Similarly, by assuming $r_{x_{2_h}^2,\, \overline{x'_{1_h}}} = 0$, where h designates a row, one may show that

$$\sum_{i=1}^{N} x_{i1}\, x_{i2}^2 = N\, r_{12}\, \sigma_{x_1}\, \sigma_{x_2}^2\, \alpha''_3 . \tag{28}$$

We further note that $\sigma_{x_1} = \sigma_{x_2}$ and that $\sum x_{i1}^3 = \sum x_{i2}^3 = N\, \alpha'_3\, \sigma_{x_1}^3$. Equation (15) can now be rewritten as

$$\sum x_i^3 = 2N\, \alpha'_3\, \sigma_{x_1}^3\, (1 + 3r_{12}) . \tag{29}$$

Substituting for r_{12} from (11) and dividing through by $N\, \sigma_x^3$ or its equivalent as given in (13), we find after simplification that

$$\alpha_3 = \alpha'_3 \frac{(1+R)\sqrt{2-R}}{2}, \tag{30}$$

or

$$\alpha'_3 = \alpha_3 \frac{2}{(1+R)\sqrt{2-R}}. \tag{31}$$

Relation of β_2 to β'_2 and β''_2. Raising both sides of (4) to the fourth power, we have

$$\sum x^4 = \sum x_1^4 + 4\sum x_1^3 x_2 + 6\sum x_1^2 x_2^2 + 4\sum x_1 x_2^3 + \sum x_2^4.$$

It has been pointed out previously that $\sum x_1^4 = \sum x_2^4$. The relationship

$$\sum x_1^2 x_2^2 = N\,\sigma_{x_1}^2\,\sigma_{x_2}^2\,[1 + r_{12}^2(\beta'_2 - 1)] \tag{32}$$

was originally stated (though in a different notation) by Karl Pearson (3, 4) and by Isserlis (2, 188) as holding under assumptions of rectilinear regression and homoscedasticity of both variables. The relationship

$$\sum x_1^3 x_2 = N\,\sigma_{x_1}^3\,\sigma_{x_2}\,r_{12}\,\beta'_2 \tag{33}$$

or

$$\sum x_1 x_2^3 = N\,\sigma_{x_1}\,\sigma_{x_2}^3\,r_{12}\,\beta''_2 \tag{34}$$

was presented (in a different notation) by Isserlis (2, 188), and was derived by making the assumption of rectilinear regression.

However, equations (32), (33), and (34) can be derived using far less stringent assumptions. In the notation for Figure 1, $\sum_{i=1}^{N} x_{i1}^2\, x_{i2}^2$ becomes

$$\sum_{g=1}^{G} x_{1_g}^2 \sum_{i=1}^{N_g} x_{2_{ig}}^2 . \tag{35}$$

Upon rearranging terms in (17), squaring, and summing over i in column g, we find

$$\sum_{i=1}^{N_g} x_{2_{ig}}^2 = \sum_{i=1}^{N_g} x'^2_{2_{ig}} + 2r_{12}\frac{\sigma_{x_2}}{\sigma_{x_1}} x_{1_g} \sum_{i=1}^{N_g} x'_{2_{ig}} + N_g \frac{\sigma_{x_2}^2}{\sigma_{x_1}^2} r_{12}^2\, x_{1_g}^2 . \tag{36}$$

Substituting from (36) in (35), we may now write

$$\sum_{i=1}^{N} x_{i1}^2 x_{i2}^2 = \sum_{g=1}^{G} x_{1_g}^2 \sum_{i=1}^{N_g} x'^2_{2_{ig}}$$

$$+ 2\, r_{12} \frac{\sigma_{x_2}}{\sigma_{x_1}} \sum_{g=1}^{G} x_{1_g}^3 \sum_{i=1}^{N_g} x'_{2_{ig}} \tag{37}$$

$$+ r_{12}^2 \frac{\sigma_{x_2}^2}{\sigma_{x_1}^2} \sum_{g=1}^{G} N_g\, x_{1_g}^4 .$$

We shall now assume that the column error of estimate, $\sigma_{e_g}^2$, is not correlated with $x_{1_g}^2$, that is, that

$$r_{x_{1_g}^2,\, \sigma_{e_g}^2} = 0, \tag{38}$$

where

$$\sigma_{e_g}^2 \equiv \frac{\sum_{i=1}^{N_g} (x'_{2_{ig}})^2}{N_g}, \tag{39}$$

and

$$r_{x_{1_g}^2,\, \sigma_{e_g}^2} \equiv \frac{\sum_{g=1}^{G} N_g\, (x_{1_g}^2)\, (\sigma_{e_g}^2) - N\, M_{x_{1_g}^2}\, M_{\sigma_{e_g}^2}}{N\, \sigma_{x_{1_g}^2}\, \sigma_{\sigma_{e_g}^2}} .$$

When this correlation is zero,

$$\sum_{g=1}^{G} N_g\, (x_{1_g}^2)\, (\sigma_{e_g}^2) = N\, M_{x_{1_g}^2}\, M_{\sigma_{e_g}^2}$$

$$= N \left[\frac{\sum_{g=1}^{G} N_g\, x_{1_g}^2}{\sum_{g=1}^{G} N_g} \right] \left[\frac{\sum_{g=1}^{G} N_g\, \sigma_{e_g}^2}{\sum_{g=1}^{G} N_g} \right] . \tag{40}$$

Now

$$\sum_{g=1}^{G} N_g = N, \quad \frac{\sum_{g=1}^{G} N_g\, x_{1_g}^2}{N} = \sigma_{x_1}^2 ,$$

and

$$\frac{\sum_{g=1}^{G} N_g \, \sigma_{e_g}^2}{N} = \sigma_{x_2}^2 \, (1 - r_{12}^2) .$$

Hence

$$\sum_{g=1}^{G} N_g \, (x_{1_g}^2) \, (\sigma_{e_g}^2) = N \, \sigma_{x_1}^2 \, \sigma_{x_2}^2 \, (1 - r_{12}^2) . \tag{41}$$

We shall further assume that the mean of the errors of estimate for a column, $\overline{x'_{2_g}}$, is not correlated with $x_{1_g}^3$, that is, that

$$r_{x_{1_g}^3, \, \overline{x'_{2_g}}} = 0 , \tag{42}$$

where $\overline{x'_{2_g}}$ is defined by (22) and

$$r_{x_{1_g}^3, \, \overline{x'_{2_g}}} \equiv \frac{\sum_{g=1}^{G} N_g \, (x_{1_g}^3) \, (\overline{x'_{2_g}}) - N \, M_{x_{1_g}^3} \, M_{\overline{x'_{2_g}}}}{N \, \sigma_{x_{1_g}^3} \, \sigma_{\overline{x'_{2_g}}}} . \tag{43}$$

When this correlation is zero,

$$\begin{aligned} \sum_{g=1}^{G} N_g \, (x_{1_g}^3) \, (\overline{x'_{2_g}}) &= N \, M_{x_{1_g}^3} \, M_{\overline{x'_{2_g}}} \\ &= N \left[\frac{\sum_{g=1}^{G} N_g \, x_{1_g}^3}{\sum_{g=1}^{G} N_g} \right] \left[\frac{\sum_{g=1}^{G} N_g \, \overline{x'_{2_g}}}{\sum_{g=1}^{G} N_g} \right] . \end{aligned} \tag{44}$$

The quantity $\sum_{g=1}^{G} N_g \, \overline{x'_{2_g}}$ was previously shown to be zero. Hence with the assumption in (42), we may write

$$\sum_{g=1}^{G} N_g \, (x_{1_g}^3) \, (\overline{x'_{2_g}}) = 0 . \tag{45}$$

From (39) and (41) we can write

$$\sum_{g=1}^{G} x_{1_g}^2 \sum_{i=1}^{N_g} x'^2_{2_{ig}} = N \, \sigma_{x_1}^2 \, \sigma_{x_2}^2 \, (1 - r_{12}^2) . \tag{46}$$

From (22) and (45) we can write

$$\sum_{g=1}^{G} x_{1_g}^{3} \sum_{i=1}^{N_g} x'_{2_{ig}} = 0. \tag{47}$$

Substituting from (46) and (47) in (37), we find

$$\sum_{i=1}^{N} x_{i1}^{2}\, x_{i2}^{2} = N\, \sigma_{x_1}^{2}\, \sigma_{x_2}^{2}\, (1 - r_{12}^{2}) + r_{12}^{2}\, \frac{\sigma_{x_2}^{2}}{\sigma_{x_1}^{2}} \sum_{g=1}^{G} N_g\, x_{1_g}^{4} . \tag{48}$$

The sum involved in the last term on the right is equivalent to $N\sigma_{x_1}^{4}\, \beta_2'$. After substituting and simplifying, we find

$$\sum_{i=1}^{N} x_{i1}^{2}\, x_{i2}^{2} = N\, \sigma_{x_1}^{2}\, \sigma_{x_2}^{2}\, [1 + r_{12}^{2} (\beta_2' - 1)] , \tag{49}$$

which is the same as the result of Pearson and Isserlis, given in (32).

Now consider (33) or (34). Referring to Figure 1, we may state

$$\sum_{i=1}^{N} x_{i1}^{3}\, x_{i2} = \sum_{g=1}^{G} x_{1_g}^{3} \sum_{i=1}^{N_g} x_{2_{ig}} . \tag{50}$$

Substituting from (19) in (50), we have

$$\sum_{i=1}^{N} x_{i1}^{3}\, x_{i2} = \sum_{g=1}^{G} x_{1_g}^{3} \sum_{i=1}^{N_g} x'_{2_{ig}} + r_{12}\, \frac{\sigma_{x_2}}{\sigma_{x_1}} \sum_{g=1}^{G} N_g\, x_{1_g}^{4} . \tag{51}$$

From (47) the first term on the right can be seen to be zero. Simplifying the last term on the right as for (49), we may then write

$$\sum_{i=1}^{N} x_{i1}^{3}\, x_{i2} = N\, \sigma_{x_1}^{3}\, \sigma_{x_2}\, r_{12}\, \beta_2' , \tag{52}$$

which is the same as the result of Isserlis, given by (33). Equation (34) can be derived in a manner similar to that stated for (33), by assuming $r_{x_{2_h}^{3},\, \overline{x_{1_h}'}} = 0.$*

**Reasonableness of the Assumptions of Zero Correlations in Equations (21), (38), and (42)*: As was noted previously, it is possible to derive (49) and (52) by making the assumptions of complete rectilinearity and homoscedasticity. However, these latter assumptions not only imply the assumptions stated in (21), (38), and (42), but also a great deal more. For the purposes of the present derivation it is *not* necessary for the column (or row) means to lie on the regression line, and for the column (or row) variances all to be identical. It is necessary only to assume that there is no consistent trend in the values of the column (or row) errors of estimate as would be represented by a correlation of the mean of the errors of estimate in a column (or row) with the square and

If we now note that $\sum x_1^4 = N\,\sigma_{x_1}^4\,\beta_2'$ and $\sum x_2^4 = N\,\sigma_{x_2}^4\,\beta_2''$, and if these equivalents together with the values for $\sum x_1^3\,x_2$, $\sum x_1\,x_2^3$, and $\sum x_1^2\,x_2^2$ given in (33), (34), and (32) are substituted in the expression for $\sum x^4$ given at the beginning of this subsection, and the resulting expression is divided through by $N\,\sigma_x^4$ and then simplified, we find that

$$\frac{\sum x^4}{N\sigma_x^4} = \beta_2 = \frac{2\,N\,\sigma_{x_1}^4\,[\beta_2' + 4r_{12}\beta_2' + 3 + 3\beta_2'\,r_{12}^2 - 3r_{12}^2]}{N\,\sigma_x^4}. \tag{53}$$

After substituting for $\sigma_{x_1}^2$ from (14) and for r_{12} from (11), and then simplifying, it is found that

$$\beta_2 = \beta_2'\,\frac{(1+R)}{2} + \frac{3(1-R)}{2} \tag{54}$$

and

$$\beta_2' = \beta_2\,\frac{2}{(1+R)} - \frac{3(1-R)}{(1+R)}. \tag{55}$$

$\sum xy$:

$$\begin{aligned}\sum xy &= \sum (x_1 + x_2)\,(x_1 - x_2)^2 \\ &= \sum x_1^3 - \sum x_1^2\,x_2 - \sum x_1\,x_2^2 + \sum x_2^3.\end{aligned} \tag{56}$$

By use of the relation $\sum x_1^3 = \sum x_2^3$, together with the expressions for $\sum x_1^2\,x_2$ and $\sum x_1\,x_2^2$ previously developed, (56) can be restated as

$$\sum xy = 2\sum x_1^3\,(1 - r_{12}). \tag{57}$$

Now $\sum x_1^3$ is equivalent to $N\,\sigma_{x_1}^3\,\alpha_3'$; and if this value and the value of r_{12} as given in (11) together with the value of α_3' as given in (31) are substituted in (57) and the expression is then simplified, it is found that

$$\sum xy = N\,\sigma_x^3\,\alpha_3\,\frac{(1-R)}{(1+R)}. \tag{58}$$

$\sum x^2 y$:

$$\begin{aligned}\sum x^2 y &= \sum (x_1 + x_2)^2\,(x_1 - x_2)^2 \\ &= \sum x_1^4 - 2\sum x_1^2\,x_2^2 + \sum x_2^4.\end{aligned} \tag{59}$$

---

the cube of the predictor score, and by a correlation of the column (or row) standard error of estimate with the square of the predictor test score. The only aspect of rectilinearity and homoscedasticity which needs to be assumed here is given in (21), (38), and (42) for the columns and the corresponding assumptions needed for the rows.

If for $\sum x_1^4$, $\sum x_2^4$, and $\sum x_1^2\ x_2^2$ there are substituted the equivalent values given previously, and it is further noted that $\sigma_{x_1} = \sigma_{x_2}$, (59) can be written

$$\sum x^2 y = 2N\beta_2'\ \sigma_{x_1}^4 - 2N\ \sigma_{x_1}^4\ [1 + r_{12}^2(\beta_2' - 1)]. \tag{60}$$

After substituting for σ_{x_1} from (14) and for r_{12} from (11), we find after simplification that

$$\sum x^2 y = \frac{N\ \sigma_x^4}{2}\ [(1-R)\,(\beta_2' - 1)]. \tag{61}$$

If the equivalent in terms of β_2 as given in (55) is now substituted for β_2', the resulting expression quickly reduces to

$$\sum x^2 y = N\ \sigma_x^4 \frac{(1-R)}{(1+R)}\ (\beta_2 - 2 + R). \tag{62}$$

$\sum y$: y has been defined as the square of the difference between a given individual's scores on two parallel tests. Hence we may write

$$\begin{aligned}\sum y &= \sum (x_1 - x_2)^2 \\ &= \sum x_1^2 - 2\sum x_1 x_2 + \sum x_2^2 \\ &= N\ \sigma_{x_1}^2 - 2\ N\ \sigma_{x_1}\ \sigma_{x_2}\ r_{12}\ + N\sigma_{x_2}^2 .\end{aligned}$$

Substituting for σ_{x_1} from (14) and for r_{12} from (11), and keeping in mind that $\sigma_{x_1} = \sigma_{x_2}$, we find that

$$\sum y = N\ \sigma_x^2 (1-R). \tag{63}$$

Solution for Parameters

It is now possible to return to the determinantal solutions for a, b, and c, and derive expressions for these parameters on the basis of the assumptions which have been made and the preliminary derivations which have been carried out.

Since x is a deviation score, $\sum x = 0$. The determinant for a is then equivalent to

$$a = \frac{N\sum x^2 y \sum x^2 - \sum y\,(\sum x^2)^2 - N\sum xy \sum x^3}{N\sum x^4 \sum x^2 - (\sum x^2)^3 - N\ (\sum x^3)^2}. \tag{64}$$

Substituting for $\sum x^2 y$ from (62); $N\sigma_x^2$ for $\sum x^2$; for $\sum y$ from (63); for $\sum xy$ from (58); $\beta_2 N\sigma_x^4$ for $\sum x^4$; and $N\ \alpha_3\ \sigma_x^3$ for $\sum x^3$, and then dividing both numerator and denominator of the right-hand side of the expression by $N^3\ \sigma_x^6$, we find after simplification that

$$a = \frac{(1-R)\,(\beta_2 - 3 - \alpha_3^2)}{(1+R)\,(\beta_2 - 1 - \alpha_3^2)}. \tag{65}$$

The determinant for b is equivalent to

$$b = \frac{N\sum x^4 \sum xy + \sum x^2 \sum x^3 \sum y - \sum xy\,(\sum x^2)^2 - N\sum x^3 \sum x^2 y}{N\sum x^4 \sum x^2 - (\sum x^2)^3 - N(\sum x^3)^2}. \tag{66}$$

Substituting $\beta_2 N \sigma_x^4$ for $\sum x^4$; for $\sum xy$ from (58); $N\alpha_3\,\sigma_x^3$ for $\sum x^3$; for $\sum y$ from (63); $N\sigma_x^2$ for $\sum x^2$; and for $\sum x^2 y$ from (62); dividing through the numerator by $N^3\,\sigma_x^7$ and the denominator by $N^3\,\sigma_x^6$, and then cancelling, we find after simplifying that

$$b = \frac{2\,\sigma_x\,\alpha_3\,(1-R)}{(1+R)\,(\beta_2 - 1 - \alpha_3^2)}. \tag{67}$$

The determinant for c is equivalent to

$$c = \frac{\sum x^4 \sum x^2 \sum y + \sum x^3 \sum xy \sum x^2 - (\sum x^2)^2 \sum x^2 y - \sum y\,(\sum x^3)^2}{N\sum x^4 \sum x^2 - (\sum x^2)^3 - N(\sum x^3)^2}. \tag{68}$$

Substituting $N\beta_2\,\sigma_x^4$ for $\sum x^4$; $N\sigma_x^2$ for $\sum x^2$; for $\sum y$ from (63); $N\alpha_3\sigma_x^3$ for $\sum x^3$; for $\sum xy$ from (58); for $\sum x^2 y$ from (62); and then dividing through the numerator by $N^3\,\sigma_x^8$ and through the denominator by $N^3\,\sigma_x^6$, and cancelling, we find after simplifying that

$$c = \frac{\sigma_x^2 (1-R)\,(\beta_2 R - \alpha_3^2 R + 2 - R)}{(1+R)\,(\beta_2 - 1 - \alpha_3^2)}. \tag{69}$$

It is now possible to rewrite equation (10) substituting for a, b, and c the expressions for these parameters given in (65), (67), and (69). The resulting equation is

$$\begin{aligned} \dot{y} = {} & \frac{(1-R)\,(\beta_2 - 3 - \alpha_3^2)}{(1+R)\,(\beta_2 - 1 - \alpha_3^2)}\,x^2 + \frac{2\,\sigma_x\,\alpha_3\,(1-R)}{(1+R)\,(\beta_2 - 1 - \alpha_3^2)}\,x \\ & \frac{\sigma_x^2 (1-R)\,(\beta_2 R - \alpha_3^2 R + 2 - R)}{(1+R)\,(\beta_2 - 1 - \alpha_3^2)}. \end{aligned} \tag{70}$$

This can be factored to some extent and then expressed as

$$\begin{aligned} \dot{y} = {} & \frac{(1-R)}{(1+R)\,(\beta_2 - 1 - \alpha_3^2)} \{ (\beta_2 - 3 - \alpha_3^2) x^2 + 2\sigma_x \alpha_3\, x \\ & + (\beta_2 R - \alpha_3^2 + 2 - R)\,\sigma_x^2 \}. \end{aligned} \tag{71}$$

It may be of interest to express three special cases in equation

form. For the case of a symmetrical distribution of x, i.e., zero skewness, $\alpha_3 = 0$. Equation (71) then reduces to

$$\dot{y} = \frac{(1-R)}{(1+R)(\beta_2 - 1)} \{(\beta_2 - 3)x^2 + (\beta_2 R + 2 - R)\sigma_x^2\}. \qquad (72)$$

For the case of kurtosis equal to that of the normal curve, i.e., $\beta_2 = 3$, (71) reduces to

$$\dot{y} = \frac{(1-R)}{(1+R)(2-\alpha_3^2)} \{-\alpha_3^2 x^2 + 2\,\sigma_x \alpha_3 x + (2R - \alpha_3^2 + 2)\sigma_x^2\}. \qquad (73)$$

For the case of zero skewness ($\alpha_3 = 0$) and kurtosis of 3, equation (71) reduces to

$$\dot{y} = (1-R)\sigma_x^2. \qquad (74)$$

The right-hand member of (74) can readily be recognized as the square of the usual expression for the standard error of measurement.

Special Points to be Noted: By successive differentiations it can be shown that the critical point, x_c, of equation (71) is

$$x_c = -\frac{\sigma_x \alpha_3}{(\beta_2 - 3 - \alpha_3^2)}, \qquad (75)$$

and that

$$\frac{d^2 \dot{y}}{d x^2} = 2\,\frac{(1-R)}{(1+R)}\,\frac{(\beta_2 - 3 - \alpha_3^2)}{(\beta_2 - 1 - \alpha_3^2)}. \qquad (76)$$

J. E. Wilkins (5, 334) has demonstrated that $\beta_2 \geqq \alpha_3^2 + 1$. Since R is a non-negative quantity, it is then evident that the sign of the second derivative depends on the sign of $(\beta_2 - 3 - \alpha_3^2)$.

When $\beta_2 - 3 > \alpha_3^2$, x_c is *negative* for positive skewness ($\alpha_3 > 0$) and *positive* for negative skewness ($\alpha_3 < 0$). From (76) it follows that these critical values are *minima*.

When $\beta_2 - 3 < \alpha_3^2$, x_c is *positive* for positive skewness ($\alpha_3 > 0$) and *negative* for negative skewness ($\alpha_3 < 0$). From (76) it follows that these critical values are *maxima*.

When $\beta_2 - 3 = \alpha_3^2$, the coefficient of x^2 in (71) is zero, and a linear relationship exists between the square of the standard error of measurement and the total test score.

EMPIRICAL VERIFICATION

To test the adequacy with which the derived general equation developed in the previous section described the manner in which the standard error of measurement varied as the total test score changed, several attempts at empirical verification were carried out.

To provide a wide variety of types of test-score distributions, it was believed desirable to put the equation to test in nine separate cases, three degrees along a scale of skewness (positive, zero, and negative) each being combined with three degrees along a scale of kurtosis (platy-, meso-, and lepto-kurtosis). These cases were numbered according to the scheme indicated below.

			SKEWNESS	
		Negative	Zero	Positive
KURTOSIS	Lepto-kurtosis	Case 1	Case 4	Case 7
	Meso-kurtosis	Case 2	Case 5	Case 8
	Platy-kurtosis	Case 3	Case 6	Case 9

Nine tests having these various shapes of distribution curve were created synthetically.* The Educational Testing Service had available several thousand answer sheets for an examination containing 256 items on general information. This was a sufficient fund of items so that sets of questions might be chosen from it to yield the desired characteristics in the total-score distributions.

From the larger number available, 1000 answer sheets were chosen on these bases: (a) every person must have attempted every item in the section; and (b) a wide range of scores on the section should exist in the sample chosen. Since the responses to these questions were in the form of numbers written in by the subjects, it was necessary to convert the responses into marks on I.B.M. answer sheets to permit the necessary statistical measures to be secured in an economical fashion. In making the I.B.M. answer sheets, one mark was made for each item, the mark being placed in the first of a pair of columns for a correct answer and in the second of the pair for a noncorrect answer. To check the accuracy of the transfer and the adequacy of the marks for scoring purposes, the I.B.M. sheets were next scored; and in instances in which the machine score disagreed with

*Note from Table 3 that the degree to which the tests fit the specifications varied greatly. Cases 3, 6, and 9, for example, were *not* alike in kurtosis.

the hand score on the original answer sheet, the papers were carefully re-examined, and corrections made or marks re-blackened.

Item Analysis of Total Test: Using the Graphic Item Counter attachment to the I.B.M. Test-Scoring Machine, an item analysis of the 256-item section was carried out. Papers were first sorted on the basis of total score into 21 unequally-sized groups, using a constant interval along the test-score range into which to sort the groups. Counts were made separately of the number of persons in each group who marked the right and who marked the wrong answer to each item. In any case in which these did not sum to the number of persons in the group, the error was investigated and corrected.

Summing the number of persons marking the right answer in the several groups gave the total number of persons in the whole sample who had correctly responded to the item. Finding the item difficulty (defined as proportion right) was then simply a matter of moving the decimal point three places to the left, for an N of 1,000.

In finding item-test correlations, each of the cases in each of the different score intervals was treated as if it fell at the mid-point of its respective interval. (With as many as 21 score intervals, this assumption sacrifices a negligible amount of accuracy.) Mean scores on the total test for successful and for unsuccessful candidates were calculated for each item. From these, together with the proportion answering the question correctly and the standard deviation of scores on the total test, it was possible to calculate a point-biserial coefficient of correlation between item score and total test score.

Selection of Items for Synthetic Tests: A preliminary investigation had been previously carried out for the purpose of developing and testing several hypotheses as to how items should be selected in order to secure total-test-score distributions of various shapes. In this early study, a sample of 400 answer sheets loaned by the U.S. Air Force Aviation Psychology Program was employed. In this study the belief was borne out that skewness could be controlled largely by the variation of mean item difficulty. However, since easy items tended to have higher correlations with the total score than did difficult items, control on mean difficulty alone was found not to be sufficient when the attempt was being made to build a test having a symmetrical score distribution. To avoid skewness, control on item-test correlation must also be exercised.

A question about which the literature yielded no hints was that of how kurtosis might be controlled by means of item selection. It was found that a set of items of the same type all of difficulty close to .50 yielded scores with a definitely flat distribution. To secure a leptokur-

tic distribution the suggestion of using two groups of items, one on each side of .50 difficulty (for example, at .40 and .60), was investigated. It was found in this preliminary work that leptokurtosis could be secured by use of two groups of items, each quite homogeneous in difficulty, with the mean difficulty of one set at about .20 and that of the other at about .80.

To facilitate the item-selection process and permit simultaneous control over item difficulty and item-test correlation, a large two-way scatter plot of the data on the 256 items was made, using difficulty as one axis and item-test correlation as the other. Using this plot, sets of items were chosen with the intention that scores on these should be distributed in certain desired ways.*

Once a set of items had been chosen, I.B.M. keys were prepared to permit rescoring on the chosen items only, and the 1000 papers were rescored. To determine whether the resulting scores formed the desired shapes of distribution curve, the scores were tallied in step intervals and moments calculated before going on to further work. In certain instances it proved somewhat difficult to secure an acceptably close approximation to the skewness and kurtosis being sought. For example, four attempts were made in arriving at a set of items that would yield scores forming a symmetrical distribution with a kurtosis of 3, the set adopted for use in the study having a β_1 of .01 and a β_2 of 2.93 when calculated from scores grouped in step intervals.

Further Item Analyses; Creation of the Parallel Tests: After it was decided that a given set of items was acceptable in the aforementioned sense, an item-analysis was carried out for the set, for the purpose of determining the correlation between each item and the score on the particular set of items. A procedure similar to the one used for the item-analysis of the 256-item test was employed. A two-way scatter plot of item difficulty versus item-synthetic test correlation was next made in each case. Each of these synthetic tests was then divided into two parallel tests through careful pairings of items on the correct scatter-plot. In making these pairings, due regard was given to the particular items involved, it being deemed desirable to match wherever possible not only on the statistical bases but also on item content and placement in the original test.

Special scoring keys were next punched to permit the obtaining of the scores on the two parallel tests on a single insertion of the answer sheet in the test-scoring machine. A rights key punched in the usual fashion provided one of these scores, the other being ob-

*See Table 5 for further data on item selection.

tained by use of an elimination key in which *all but* the correct responses to the other test were punched out. Two checks on the scoring were used: (1) the sum of the two scores thus obtained should equal that of the total synthetic test; and (2) a rescoring of one parallel test was made to check whether the two scores had been recorded in a consistent arrangement. (It may be appropriate to mention here that all scoring on the I. B. M. machines was checked by rerunning the papers through the machine, with discrepancies being cared for by hand scoring and examination of the answer sheets for possible sources of error.)

Scatter Plots and Further Statistical Treatment: Once the scores had been obtained on pairs of parallel tests, unit-interval scatter plots were prepared of score on one parallel test against score on its mate. These scatter plots provided the basis for calculating the Pearson product-moment coefficients of correlation between the scores on the half-tests or parallel forms. Table 1 presents these correlations, together with reliabilities of the synthetic tests as estimated by use of the Spearman-Brown formula. From these scatter plots there were also calculated the means, standard deviations, and measures of skewness and kurtosis of the half-tests or parallel forms. These data are presented in Table 2.

By examination of Table 2 one can judge the effectiveness of the procedure used to create the parallel tests, and also judge whether the assumptions made in the theoretical development correspond in a reasonable fashion to existing parallel forms built on the basis of the best techniques feasible at present. It can be noted that the means and the standard deviations agree very well. Slightly more variation is observed in α_3, β_1, and β_2, but nevertheless the matching seems satisfactorily close even on these measures, except for Case 3. (Note $\beta_1 = \alpha_3^2$.)

From the scatter-plot of half-score against half-score there was derived a plot in which the total score on the synthetic test was plotted against the signed difference between scores on the half tests. This plot necessarily has a hexagonal checkerboard character, for with a given total score only certain possible differences exist. (With a total score of 3, for example, the possible differences are 3, 1, —1, and —3.) Next, this plot was used to create still another plot, in which total score on the synthetic test was plotted against the absolute (unsigned) difference between scores on the parallel forms that together constituted the synthetic test. Again a checkerboard pattern is obtained; furthermore, because only certain differences can occur, this array has the appearance of an isosceles trapezoid. The

sloping at the ends is due to what it will be convenient to designate as the *end effect.* (This phenomenon is discussed further on page 212.)

Curve-Fitting and Graphs: From this final plot the quantities $\sum x^2$, $\sum x^3$, $\sum x^4$, $\sum y$, $\sum xy$, and $\sum x^2y$ were derived. From these quantities were obtained the coefficients of the second-degree curve of best fit between the total score on the synthetic test as the independent variable and the square of the difference between the scores on the two half-tests as the dependent variable. Also from them were obtained the standard deviation, α_3, β_1, and β_2 for the total-test-score distribution. These quantities together with the reliability of the synthetic test were substituted in equation (71) to secure the theoretical curve of relationship between standard error of measurement and total test score. The values of the quantities, together with the test means, are presented in Table 3.

The second-degree curve of best fit, which henceforth will be termed the empirical curve, and the derived curve were in terms of deviation scores on the test variable. For ease of plotting it was convenient to carry out a simple transformation to re-express these curves in terms of raw scores on the test. The coefficients of these curves are presented in Table 4.

The means of the squares of differences between parallel-test scores were calculated at each unit of score along the total synthetic-test-score scale. Also, to secure a somewhat more stable value, the means of the squared differences were found for units of five score points along the total-score axis.

The empirical and derived curves were then drawn in the same coordinate system, and the column means and grouped-column means plotted on this same graph. In addition, a line was drawn in parallel to the x-axis to represent the prevailing assumption that the standard error of measurement is constant throughout the test-score range. The set of nine graphs corresponding to the nine cases studied in the empirical investigation are presented as Figures 2 through 10.

Evaluation

For judging the adequacy of the derived general equation several criteria were thus provided in the graphs.* (It should be pointed out that if the assumptions made in the derivation were to be perfectly true for each actual pair of parallel tests, the derived and empirical curves would coincide.) The most basic indication of the mag-

*The advisability of applying tests of statistical significance was discussed with Professor S. S. Wilks, who indicated it would be inappropriate to apply such a test to these data. Professor Wilks favored the procedure used.

nitude of the standard error of measurement actually observed at the several points in the test-score distribution is provided by the means of the squared differences for individual or groups of columns. In examining Figures 2 through 10 critically, therefore, one should not merely ask how closely the derived curve follows the empirical, but one should also evaluate the derived curve as to the accuracy with which it portrays the variation shown in the individual column means and grouped column means.

To secure some quantitative estimate of the goodness of the fit of the derived curve to the data, the procedure was adopted of computing the differences between the value of the derived curve and that of each of the grouped-column means. For purposes of comparison and evaluation, a similar procedure was carried through for the empirical curve and for the zero-slope straight line which represents the assumption of a constant standard error of measurement. This method assumed that all grouped-column means were equally stable,* when of course those near the mode were based on many more scores than those a considerable distance away from the mode. However, since the main interest was in the over-all description of the data afforded by the curve or the straight line, the approach nevertheless was worth while, and proved definitely helpful in making evaluations, especially when considered in conjunction with the examination of Figures 2 through 10. The algebraic (signed) sum and the absolute sum of the differences are presented for each case in Table 5.

In considering the data of Table 5, one should keep in mind the significance of the signed sum and of the absolute sum. A small signed sum arises when the differences in one direction are about equal to those in the other direction, regardless of the size of these differences. A large signed sum will occur when the curve systematically misses the grouped-column means, i.e., lies mostly above or below these means. The absolute sum is small when the "misses" are small, and large when the differences are great. Judging from the data of Table 5, the derived curve provides a better description of the data for cases 1 (excluding the lowest 5% of the scores), 4, 7, 8, and 9; the derived general curve and the straight line are about equally good in describing the data for cases 2, 5, and 6; while the straight line seems a somewhat better description of the data for case 1 (when all scores are considered) and for case 3. (However, for case 3, as has been noted previously, the matching of the two parallel tests on skewness and kurtosis was less close than would have been desirable,

*If differences were weighted, the conclusions would be the same. Note that the empirical curve is the best fit to the *weighted* data.

and this makes the interpretation of the results for case 3 distinctly unclear.)

Examination of Figures 2 through 10 was also made for the purpose of noting whether there were consistent trends present in the observed error data, and for comparing the derived curve with the empirical curve of best fit. Consistent trends were noted for cases 1, 4, 6, 7, 8, and 9, and were in each case curvilinear. Obviously the zero-slope straight line representing the assumption of a constant standard error of measurement could not follow these trends. However, these trends are quite accurately described by the derived curve for cases 1 (when the lowest 5% of scores are excepted), 4, 7, 8, and 9; while for case 6 the curvilinearity of the derived curve is somewhat over-great as compared with the trend in the observed data. When the derived curve was compared with the empirical second-degree curve of best fit, good agreement was similarly noted for all cases except for case 3 and the very low score range of case 1.

General Points: The reader may raise the question as to how often the types of test-score distributions synthesized as cases 1-9 might occur in actual work with tests. Negative skewness with varying degrees of kurtosis (cases 1, 2, and 3) would occur in tests built on the mastery principle. More usual would be the symmetrical (cases 4, 5, and 6) or positively skewed distributions (cases 7, 8, and 9). High kurtosis with symmetry seems very difficult to achieve; tests with score distributions like that of case 4 probably are rarely built. Positively skewed score distributions are frequently sought in selection testing where only a small fraction of the candidates are to be chosen, as, for example, in selecting persons to be awarded scholarships.

Case 4 was noteworthy in that it was the only instance in which upturned parabolae were obtained as the empirical and derived curves. It was the single case in which the quantity ($\beta_2 - 3 - \alpha_3^2$) was positive, this being the necessary condition for the coefficient of the x^2 term in the derived equation to be positive.

For case 5 the derived curve almost coincided with the empirical curve for more than two-thirds of the range of scores. Case 5 involved the longest synthetic test used in the study. (It included 134 items.) With the over-all standard error of measurement less than 4 score points, the tails of the actual error-total score scatter plot were thus practically free from the end effect discussed below. It should be noted that if a perfectly symmetrical distribution with a kurtosis of 3 had been synthesized, the derived equation would have reduced to the zero-slope straight line.

The synthetic test for case 6 was built of items with proportions right varying from .34 to .64 with a mean of .50. A number of years ago Mrs. Thurstone stated (4, 341) that for *general* validity, tests should have this sort of item-difficulty distribution. On the other hand, tests for cases 7, 8, and 9 were constructed from items with proportions right averaging less than .50, in order to secure positive skewness in the score distributions. Frequency distributions of the proportions getting items right for each of the nine tests are given in Table 6.

End Effect: The necessary depressing effect which arises at the very end of a score distribution, where large differences between the half-scores cannot possibly occur, is termed the *end effect*. (A perfect score, for example, can be divided only one way, i.e., into two equal scores, since a person with a perfect score on the whole test necessarily has a perfect score on each half.) On each of the figures for cases 1-9 there was drawn in a step-function at the ends of the score distribution, in order to provide a graphical representation of the limiting of the largest-sized squares of differences that might occur. For the skewed cases the end-effect is especially important; small empirically observed errors of measurement are inevitable in the tail where the pile-up of scores occurs. The individual column means and grouped column means were intentionally not represented in the figures for that portion of the score range where it was believed the end effect was seriously limiting the possible magnitude of the obtained standard error of measurement. Because a very large proportion of all differences between the parallel-test scores were less than six points, and practically all less than nine points, the distance along the total score base line for which no means were plotted was usually of the order of five to eight points.

Further Points: The analytical development carried through in the present study assumed that the variation of the standard error of measurement could be adequately described by a curve of no higher than the second degree. Examination of the data presented in Figures 2 to 10 revealed no evidence that the assumption was incorrect, that is to say, that a curve of the fourth degree (for example) would be required for the purpose.

One additional point may be noted. In the type of item used in the present study guessing was an unimportant factor, for the probability of a chance correct response was very small, being of the order of 1 in 20 or less.

CONCLUSIONS

The results of the present study indicate that the assumption that the standard error of measurement is a constant appears to be tenable only under certain special conditions. The theoretical analysis showed that a constant standard error would prevail provided the test-score distribution were to be symmetrical and mesokurtic. Case 5 of this study was a close approximation to these conditions, and for case 5 a straight line fitted the empirical data quite well. In case 2 a small degree of skewness occurred combined with a kurtosis close to 3, and here also the straight line parallel to the test-score axis seemed to describe the individual and grouped column means of squared differences (i.e., the standard error of measurement) rather adequately. In none of the other cases, with the possible exception of case 4, did the results obtained bear out the hypothesis of a constant standard error of measurement. Examination of the values of β_2 and α_3^2 listed in Table 3 revealed that α_3^2 for cases 2, 4, and 5 was small, and that β_2 was close to 3.0. Hence the empirical results were in definite agreement with the theoretical analysis as to the special conditions under which the standard error of measurement would be constant.

The derived second-degree curve in which the parameters were functions of the standard deviation, skewness, kurtosis, and reliability was found to describe accurately the variation of the standard error of measurement as empirically observed at numerous points in the total test-score distribution. The only appreciable amount of divergence between theory and fact occurred in the tail of a test-score distribution having a very great skewness and kurtosis. The derived curve in each case was compared with a parabola fitted to the test score-error data by the method of least squares, and except for case 3 and the low-score range of case 1, the agreement between the two curves was found to be good. The theoretical curve was also judged by the method of computing both the differences between the grouped-column means and the values of the theoretical curve and the differences between these means and the straight-line value. Two aspects of these differences were considered: (1) the signed or algebraic sum, and (2) the absolute sum. As judged by this method, the derived curve provided a representation of the data which was distinctly superior to that of the straight line in cases 1 (excluding the lowest 5% of the scores), 4, 7, 8, and 9. For cases 2, 5, and 6 the derived curve and the zero-slope straight line were about equally good representations of the error data. For case 1 (all scores being considered)

and 3,* neither the derived curve nor the zero-slope straight line described the data satisfactorily. When both aspects of the method of computing differences were considered together with the examination of Figures 2-10, including the comparison of the theoretical curve with the empirical curve, it was concluded that the theoretical curve provided a more satisfactory representation of the variation and magnitude of the standard error of measurement than the zero-slope straight line when both of these were applied to a number of tests having widely different types of total-score distributions.

Recalling that the square of the standard error of measurement was plotted, one can state that the difference between the standard error as empirically observed and the standard error as calculated from the derived formula would for practically all scores in the nine cases be but a small fraction of a score point.

The present study showed that the theoretical curve was quite satisfactory for application to a variety of tests in which guessing was an unimportant factor in success on an item. Should further work on different tests indicate that this theoretical curve or a modification provides an adequate representation of the variation and magnitude of the standard error of measurement for tests in which guessing is a factor of importance in the success on an item (for example, a true-false test), the theoretical curve would be demonstrated to have definite value in testing programs for deriving the magnitude of the standard error of measurement at frequent intervals in the test-score range. The present findings indicate that this procedure would be preferable to the practice of stating the test reliability or the conventional over-all standard error of measurement especially in those instances in which distributions are obtained which are *not* symmetrical and mesokurtic. (Dr. Ledyard Tucker has stated to the writer that in his experience most test-score distributions are usually distinctly flat and oftentimes are skewed. The data presented in Conrad's report, (1, 21) on Navy classification tests showed that of sixteen measures, eleven were platykurtic and only one was leptokurtic. Five of these tests had significant skewness and several others had a skewness close to statistical significance. Consequently, for most tests the conventional practice in stating the standard error of measurement appears to be rather dubious.)

If the present findings are duplicated in other studies involving tests in which guessing is an important factor in success on an item, the author would then recommend that the descriptive information

*For case 3, it is to be noted from Table 2 that the match between the two halves was poorest of all nine cases. Hence in this case the assumptions made on p. 192 are *least* adequately met in the actual parallel tests.

provided with objective standardized tests include the standard error of measurement at ten or more points equally spaced on the total-score scale, calculated using the moments of the total-score distribution and the reliability estimated from the correlation between matched halves of the test.

REFERENCES

1. Conrad, H. S. A statistical evaluation of the basic classification test battery (*Form* I), *O.S.R.D. Report 4346*. Washington: Applied Psychology Panel, N.D.R.C., 1945, 93 pp.
2. Isserlis, L. On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression. *Biometrika*, 1916, 11, 185-190.
3. Pearson, K. On the probable errors of frequency constants. *Biometrika*, 1913, 9, 1-10.
4. Thurstone, T. G. The difficulty of a test and its diagnostic value. *J. educ. Psychol.*, 1932, 23, 335-343.
5. Wilkins, J. E. A note on skewness and kurtosis. *Ann. math. Stat.*, 1944, 15, 333-335.

TABLE 1

Number of Items, Correlation between Half-Scores, and Reliability Derived from Application of the Spearman-Brown Formula, for Each of Nine Synthetic Tests Used in Error of Measurement Study
($N = 1000$)

Case	Number of Items	Correlation of Half-Scores	Reliability
1	60	.8796	.9359
2	66	.8722	.9317
3	78	.9024	.9487
4	64	.8381	.9119
5	134	.8935	.9438
6	64	.8917	.9427
7	66	.8851	.9277
8	60	.8506	.9193
9	54	.8714	.9313

TABLE 2

Means, Standard Deviations, and Measures of Skewness and Kurtosis of Half-Score Distributions for Nine Synthetic Tests Used in Error of Measurement Study
($N = 1000$)

Case and Part	Mean	Standard Deviation	a_3	β_1	β_2
1, 1st Half	23.747	5.589	—1.419	2.014	4.783
1, 2nd Half	23.880	5.617	—1.404	1.971	4.797
2, 1st Half	15.762	5.352	— .247	.061	2.886
2, 2nd Half	15.787	5.349	— .286	.082	2.911
3, 1st Half	24.836	8.090	— .656	.431	2.890
3, 2nd Half	24.843	8.154	— .524	.275	2.663
4, 1st Half	16.028	4.413	— .178	.032	3.187
4, 2nd Half	16.022	4.398	— .171	.029	3.272
5, 1st Half	28.256	8.051	— .100	.010	2.891
5, 2nd Half	28.248	8.001	— .054	.003	2.877
6, 1st Half	15.964	7.323	— .078	.006	2.212
6, 2nd Half	15.991	7.317	— .034	.001	2.101
7, 1st Half	6.651	5.343	+ .849	.721	3.084
7, 2nd Half	6.672	5.285	+ .919	.845	3.329
8, 1st Half	6.757	5.116	+ .822	.675	3.067
8, 2nd Half	6.751	5.110	+ .818	.670	3.107
9, 1st Half	12.769	6.686	+ .242	.059	2.300
9, 2nd Half	12.820	6.674	+ .283	.080	2.279

TABLE 3
Means, Standard Deviations, and Measures of Skewness and Kurtosis for Total Score Distributions of Nine Synthetic Tests Used in Error of Measurement Study
($N = 1000$)

Case	Mean	Standard Deviation	α_3	β_1	β_2
1	47.627	10.864	—1.467	2.151	4.947
2	31.549	10.354	— .305	.093	2.919
3	49.679	15.842	— .623	.388	2.820
4	32.050	8.447	— .182	.033	3.171
5	56.504	15.618	— .097	.010	2.903
6	31.955	14.238	— .071	.005	2.145
7	13.323	10.264	+ .902	.813	3.212
8	13.508	9.837	+ .816	.667	3.054
9	25.589	12.923	+ .267	.071	2.278

TABLE 4
Table of Coefficients of the Empirical Second-Degree Curve of Best Fit and of the Derived Curve Each with Origin at Raw Score of Zero

Case		Coefficient a	b	c
1	Emp.	—0.015142	+0.90283	+ 0.71007
	Theor.	—0.003753	—0.22984	+27.46682
2	Emp.	—0.000622	+0.07488	+ 5.64469
	Theor.	—0.003382	+0.09096	+ 8.18082
3	Emp.	—0.009773	+0.80983	— 0.77584
	Theor.	—0.010452	+0.67521	+ 7.74452
4	Emp.	+0.007328	—0.50224	+14.33050
	Theor.	+0.002972	—0.25696	+11.25703
5	Emp.	—0.002962	+0.23675	+ 9.59395
	Theor.	—0.001625	+0.13709	+11.54532
6	Emp.	—0.012615	+0.82773	+ 0.85071
	Theor.	—0.022269	+1.37117	— 4.94411
7	Emp.	—0.014602	+0.79525	+ 1.15792
	Theor.	—0.016147	+0.92689	— 0.16597
8	Emp.	—0.015302	+0.88074	+ 0.18595
	Theor.	—0.018553	+0.98798	— 0.35534
9	Emp.	—0.013968	+0.85083	+ 0.91349
	Theor.	—0.023384	+1.40013	— 5.13937

TABLE 5

Differences between Grouped-Column Mean and Value from Derived Equation, Value from Empirical Curve, and Value of Straight Line for Nine Synthetic Test-Score Distributions

Case	Derived Equation Σd	Derived Equation $\Sigma\|d\|$	Empirical Equation Σd	Empirical Equation $\Sigma\|d\|$	Straight Line Σd	Straight Line $\Sigma\|d\|$
1 ($\alpha_3 = -1.5$, $\beta_2 = 4.9$)	+39.1	48.1	—0.3	18.5	—32.7	41.7
1 (Lowest 5% of scores excluded)	— 1.2	7.8	—0.3	7.7	—16.1	25.1
2 ($\alpha_3 = -.3$, $\beta_2 = 2.9$)	+ 1.5	16.5	+2.6	8.6	+ 4.7	10.3
3 ($\alpha_3 = -.6$, $\beta_2 = 2.8$)	+21.8	50.8	—4.4	31.8	+ 1.2	36.3
4 ($\alpha_3 = -.2$, $\beta_2 = 3.2$)	— 3.1	12.1	+2.4	11.2	— 8.2	12.8
5 ($\alpha_3 = -.1$, $\beta_2 = 2.9$)	+ 8.0	54.6	+4.0	54.8	+16.2	52.4
6 ($\alpha_3 = -.1$, $\beta_2 = 2.1$)	+ 0.3	22.5	+2.6	14.0	+ 0.9	20.9
7 ($\alpha_3 = +.9$, $\beta_2 = 3.2$)	+ 8.2	18.8	+0.6	18.8	—17.5	31.1
8 ($\alpha_3 = +.8$, $\beta_2 = 3.1$)	— 2.7	11.5	+0.1	11.5	—20.0	30.8
9 ($\alpha_3 = +.3$, $\beta_2 = 2.3$)	— 3.0	19.4	+1.4	13.4	+ 3.0	24.0

TABLE 6

Item Difficulty Indices (Proportion Right) for Items Included in the Nine Synthetic Tests

		Case 1	2	3	4	5	6	7	8	9
	.95 - .99					9				
	.90 - .94	5			10	12				
	.85 - .89	14	6		5	8				
	.80 - .84	15	15	15	8	5				
P	.75 - .79	9	7	9	5	4				
r	.70 - .74	7	2	7	4	4				
o	.65 - .69	7		7		3				
p	.60 - .64	3		2		4	8			
o	.55 - .59			13		2	16			7
r	.50 - .54			10		1	10			11
t	.45 - .49			9		3	10			10
i	.40 - .44			6		2	8			4
o	.35 - .39					5	12			3
n	.30 - .34					6		6	10	13
	.25 - .29		7		1	6		15	15	15
r	.20 - .24		9		8	4		9	9	1
i	.15 - .19		18		7	7		18	18	
g	.10 - .14		2		14	16		17	8	
h	.05 - .09				2	14		1		
t	.00 - .04					19				
	Number of Items	60	66	78	64	134	64	66	60	64

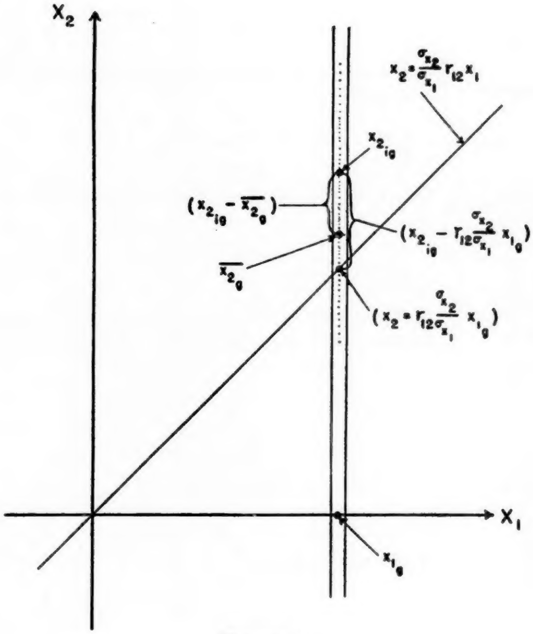

X_2
$x_2 = \frac{\sigma_{x_2}}{\sigma_{x_1}} r_{12} x_1$
$x_{2_{ig}}$
$(x_{2_{ig}} - \overline{x_{2_g}})$
$(x_{2_{ig}} - r_{12}\frac{\sigma_{x_2}}{\sigma_{x_1}} x_{1_g})$
$\overline{x_{2_g}}$
$(x_2 = r_{12}\frac{\sigma_{x_2}}{\sigma_{x_1}} x_{1_g})$
X_1
x_{1_g}

FIGURE 1

Legend for Figures 2-10

× Mean of squared differences for an individual column

● Mean of squared differences for group of five columns

Curve obtained from plotting equation (71)—Derived general curve

Second degree curve of best fit—Empirical curve

Straight line representing conventional standard error of measurement assuming this to be constant

Frequency polygon

Step function indicating end effect

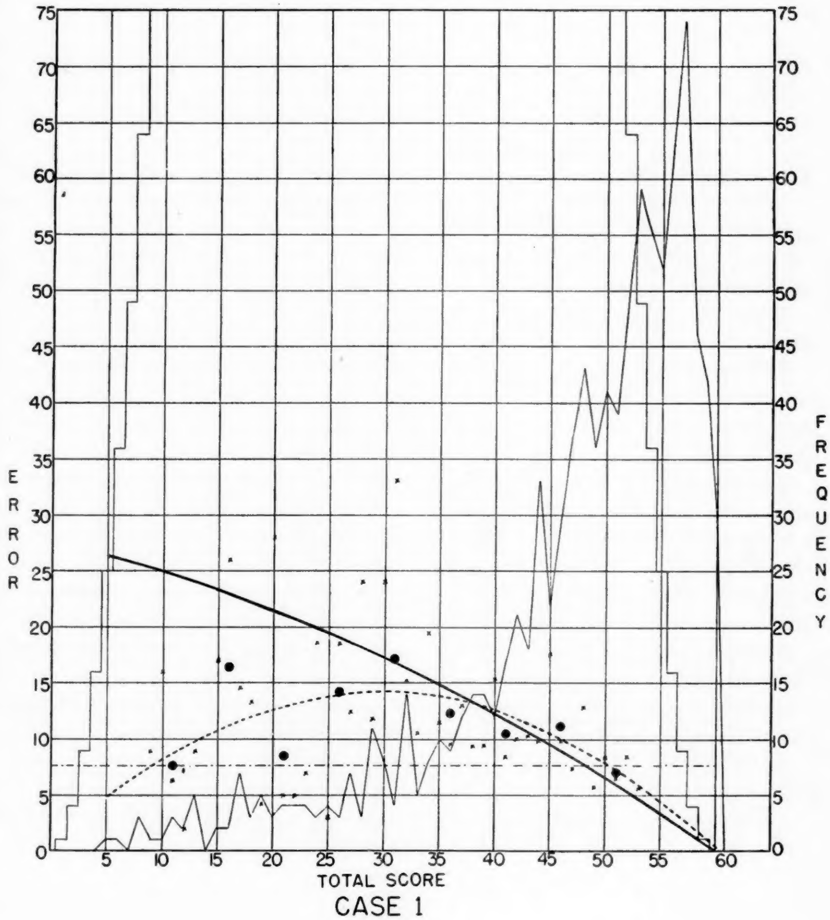
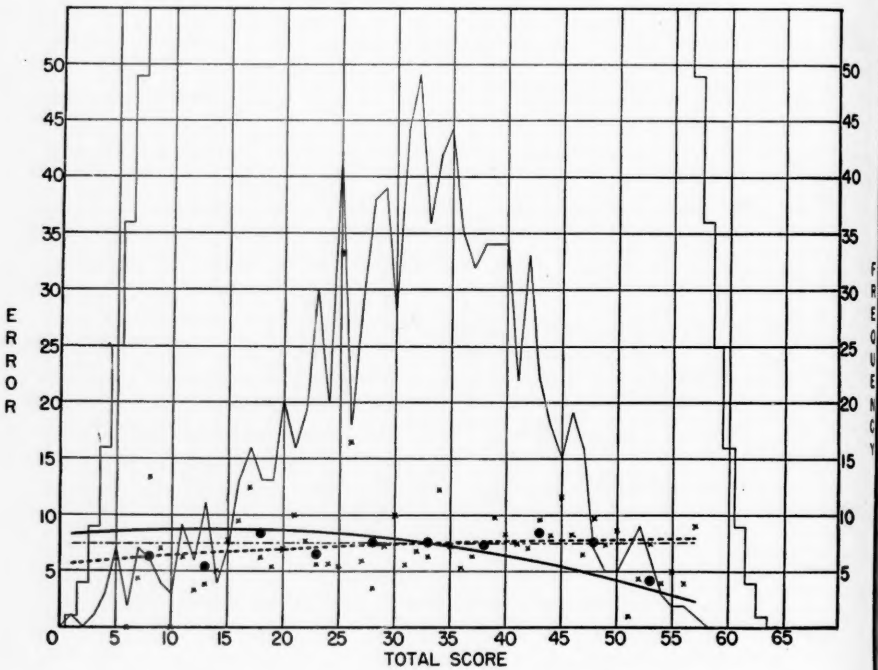
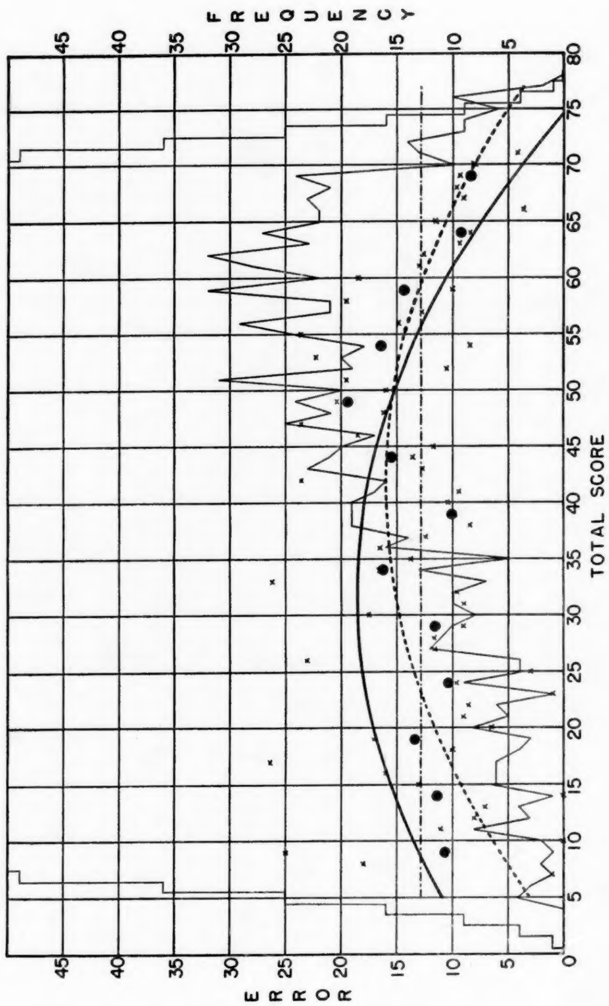


FIGURE 2



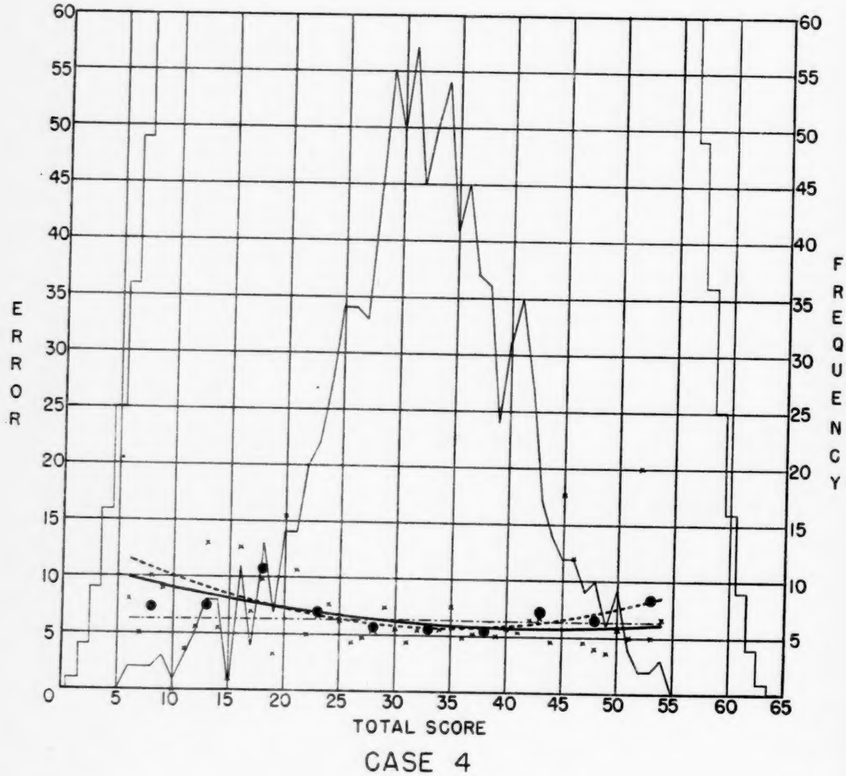
CASE 2

FIGURE 3



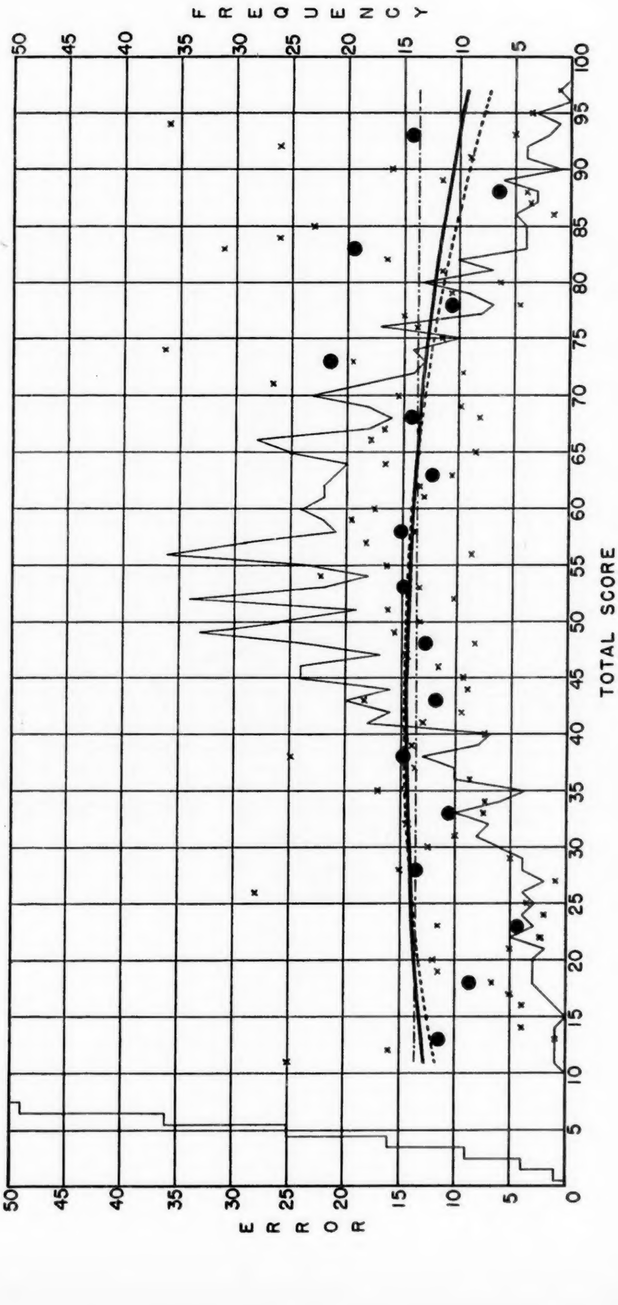
CASE 3

FIGURE 4



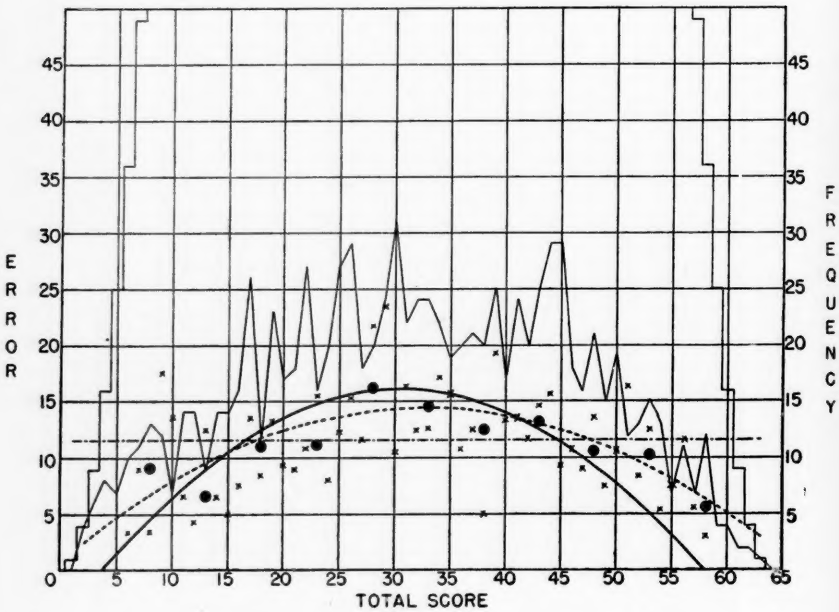
ERROR
FREQUENCY
0
5
10
15
20
25
30
35
40
45
50
55
60
65
TOTAL SCORE
CASE 4

FIGURE 5

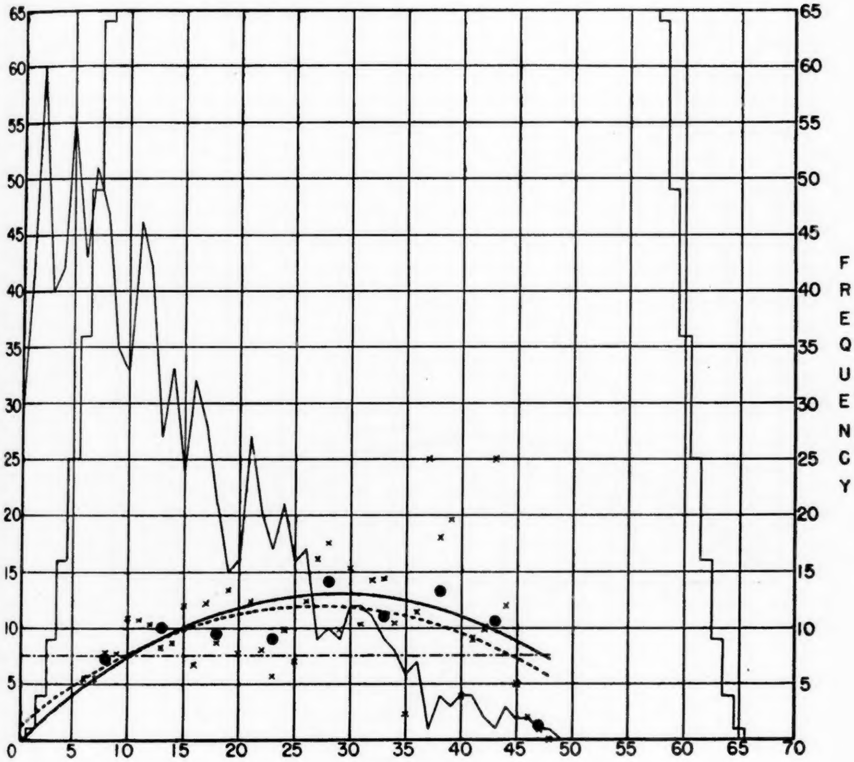


CASE 5

FIGURE 6



CASE 6
Figure 7



CASE 7

FIGURE 8

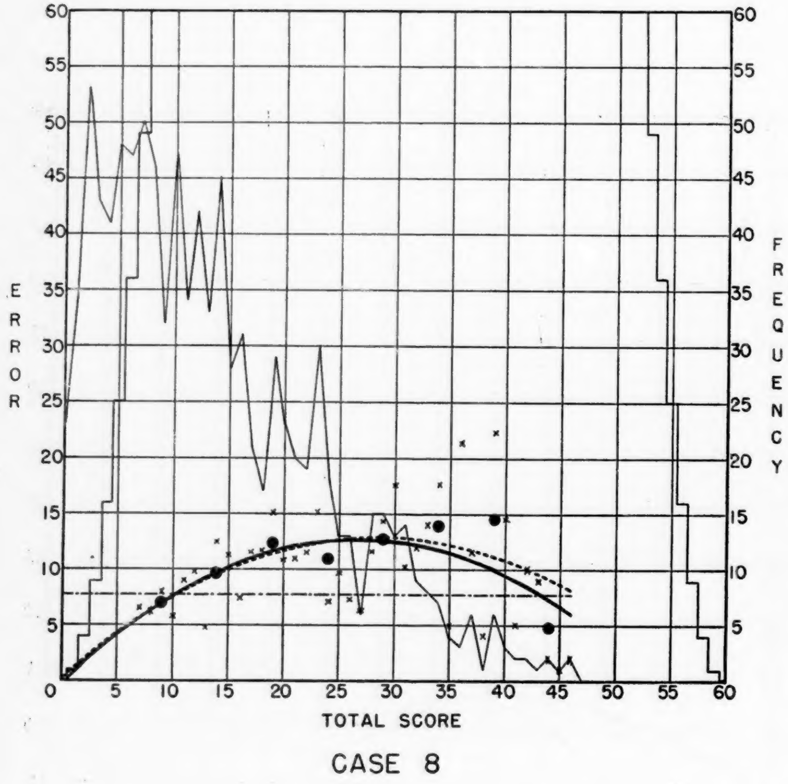
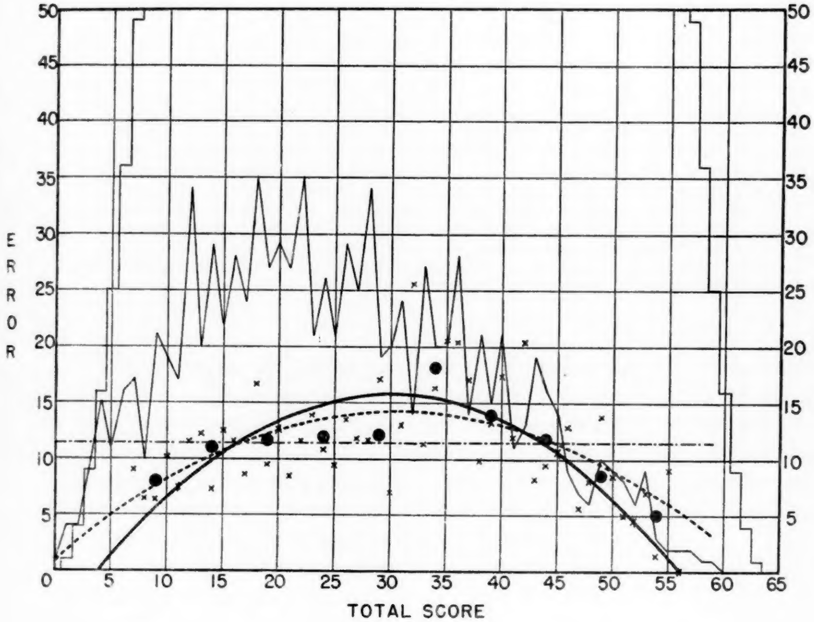

ERROR
FREQUENCY
TOTAL SCORE
0
5
10
15
20
25
30
35
40
45
50
55
60
CASE 8

FIGURE 9



CASE 9
FIGURE 10

PSYCHOMETRIKA—VOL. 14, NO. 3
SEPTEMBER, 1949

A NEW ITERATIVE METHOD FOR CORRECTING ERRONEOUS COMMUNALITY ESTIMATES IN FACTOR ANALYSIS

ROBERT J. WHERRY
THE OHIO STATE UNIVERSITY

A new method for correcting erroneous communality estimates is applicable to any completed orthogonal factor solution. It seeks, by direct correction of factor loadings, to make the residuals conform to the chance error criteria of zero mean and zero skewness for each row separately. Two numerical examples, with one and two factors, respectively, are presented. The method can be used as a short cut for Dwyer's extension in adding variables to a matrix. It can also be used as a short cut in cross-validation factor studies. Successful use on problems with many variables and numerous factors is claimed. Factors can be made oblique, *after* correction, if desired.

The purpose of factor analysis is the discovery of a meaningful minimal set of reference axes, together with projections of the involved tests upon those axes, which will yield a set of theoretical correlations differing from the original empirical correlations only by chance errors contained in the latter. This makes the residual table (containing these differences) both a test of the efficacy of the factor structure and the best cue as to its deficiency, if any.

If the residuals are due to chance errors, they should be distributed in a chance manner, i.e., (1) have a mean of zero, (2) small variability, (3) no skewness, and (4) normal kurtosis. While frequently applied to all residuals at once, these criteria can be more stringently applied to each row (or column) separately. Implications in (1) and (3), especially (1), above, as applied to each row (or column) of the residual table, lead to the formulation of the iterative process described in this paper.

In the usual Thurstone procedure, the factorial process begins with an estimation of the communality coefficients and involves a check on the accuracy of the method in that residuals at each step (including that for the communality) must add to zero and a repetition of this process for each added factor. Once the entire process is completed, an iterative procedure based upon redoing the entire process upon the basis of the obtained communalities is recommended. The entire process is repeated until the finally obtained communalities agree with those of the previous trial. While sound, this method is so exhaustive of time that it is seldom used at all and almost never

carried to exact completion. As a consequence factor studies are reported with high and skewed (by row or column) residuals and with sizable errors in the factor loadings, due to poor communality estimation. The method proposed here provides a much quicker and easier method of iteration which it is hoped will actually be used.

While better initial estimation of communalities has been investigated widely and improved considerably, it is unlikely that a perfect practicable solution will ever be reached. Until such a method is derived, the present correction method will be useful. Since many if not most investigators still use "the highest correlation coefficient in the row or column" as the basis of the "best" estimate, the correction method will be applied to errors introduced by that assumption.

Procedure

The present method can be outlined in the following steps, which will be made clearer by subsequent application to numerical examples:

1. Complete the factorial analysis by any basis of estimation of communalities and any method of factoring (centroid, group centroid, etc.), rotate to the simplest orthogonal structure, compute the residual table by use of the equation $R = FF'$, and check all computations.

2. Drop the values in the diagonals, and sum the remaining residuals for each row (or column) separately. (The final values in the diagonal should finally become zero in any case, when the iteration is complete, and therefore the sum without that value should equal zero).

3. Select the row (or column) with the highest (regardless of sign) sum of residuals to work on. Look for a pattern of residuals which follows either directly (same sign) or inversely (opposite sign) the loadings for one of the factors. Compare the magnitude of this pattern with that of the loadings and determine the amount to be added (or subtracted) to (from) the loading for that test on that factor. A rough guide to the size of this increment can be obtained by using the ratio of the sum of the residuals to the sum of loadings (omitting the loading for the variable in question), $\sum$ residuals i / ($\sum$ loadings $- l_i$). For example, if you found

	1	2	3	4	5
residuals row 3	.05	.06	—	.01	—.01
factor A	.60	.70	.40	.00	.00

you would increase the loading of .40 by + .08 [The equation, $\sum$ res. / ($\sum$ loadings $- l_i$), yields $.11/1.30 = .84$] since multiplying .60 and .70 (the other loadings) by .08 would yield .05, .06, —— , .00, and .00, which are about the size and of the same sign as the present residuals.

4. Add this increment to the loading of the test on the factor in question; multiply the increment by the remaining loadings on the factor and subtract these products from the proper cells of the residual table. Thus in the example in 3 above, you obtain:

residuals, row 3	.00	.00	—	.01	—.01
corrected factor loadings	.60	.70	.49	.00	.00

5. Get corrected sums of residuals and repeat this process until the residual table can no longer be improved. Too great accuracy in early stages is not necessary. Usually corrections should be made first in multiples of $\pm$ .10, next by multiples of either $\pm$ .05, and finally by multiples of $\pm$ 01 if that degree of accuracy is desired.

6. Recompute the residual table using the final corrected factor pattern to assure that dropping of decimals, incorrect multiplications or subtractions, etc., have not introduced computation errors into the final residuals.

7. Repeat steps 5 and 6 until no further change is indicated.

8. *A warning.* Frequently in studies involving several factors, early errors cover up or hide small factors. If the process is applied correctly, it will clear up all but a few residuals, all belonging to the same variables, which will in turn grow larger and permit the extraction of the hidden factor. When this is clearly the case, the factor should be extracted and rotated against the remaining factors (at their present stage of correction) before the process is continued.

A Single-Factor, Three-Variable Example

The writer, before presenting this example, would like to state that he is aware that in this case a much better (actually exact) method of communality estimation is available and that the single-factor case is always easiest to demonstrate. The example is given merely to show that the method does work. The single-factor case is presented first because of its simplicity.

Consider the correlation matrix,

1	2	3
1.00	.54	.36
.54	1.00	.24
.36	.24	1.00

where, using "the highest correlation" basis of estimating communality, we substitute and solve, getting:

	1	2	3		
1	*.54*	.54	.36		
2	.54	*.54*	.24		
3	.36	.24	*.36*		
	1.44	1.32	.96	3.72	1.93
	.75	.69	.50		(1.94)

The resulting residual table is

	1	2	3	Check
1	*—.022*	.022	—.015	(—.013)
2	.022	*.068*	—.105	(—.013)
3	—.015	—.105	*.110*	(—.000)

Dropping the communality entries and adding gives

	1	2	3	Σ residuals	loading	correction
1	——	.022	[—.015]	.007	[.75]	[(—.075)]
2	.022	——	[—.105]	—.083	[.69]	[(—.069)]
3	[—.015	—.105]		—.120*	.50	—.10

Row 3 has the highest sum of residuals. This sum is —.120, while the sum of factor loadings (not including test 3) is 1.44; thus a correction of approximately —.10 seems in order. Subtracting the bracketed values under the correction column from the corresponding bracketed residuals and correcting the factor loading in question yields:

	1	2	3	Σ residuals	loading	correction
1	——	[.022	.060]	.082*	.75	+.10
2	[.022]	——	—.036	—.016	[.69]	[.069]
3	[.060]	—.036	——	.024	[.40]	[.040]

Here row 1 is highest; the sum of residuals is .082, while the sum of other factor loadings is 1.090. The proper correction is +.10. Making the correction yields:

	1	2	3	Σ res.	loading	correction
1	——	[—.047]	.020	—.027	[.85]	[—.085]
2	[—.047	[——]	—.036]	—.083*	.69	—.10
3	.020	[—.036]	——	—.016	[.40]	[—.040]

Here row 2 is highest; the ratio —.083/1.250 yields —.10 to be the best correction, which yields:

	1	2	3	res.	loading	correction
1	——	[.038	.020]	.058*	.85	+.05
2	.038	——	—.004	.042	.59	+.030
3	.020	—.004	——	.016	.40	+.020

Row 1 is again highest; the ratio of .058/.990 yields —.05 as the best correction, which gives:

	1	2	3	res.	loading	correction
1	——	.008	.000	.008	[.90]	[.009]
2	.008	——	.004	.012*	.59	+.01
3	.000	.004	——	.004	[.40]	[.004]

Row 2 is largest, and the ratio of .012/1.30 yields +.01 as the best correction, which yields:

	1	2	3	res.	loading	correction
1	——	—.001	.000	—.001	.90	.00
2	—.001	——	.000	—.001	.60	.00
3	.000	.000	——	.000	.40	.00

Consideration of rows 1 and 2, which are equally high, in neither case (ratios of —.001/1.000 and —.001/1.300, respectively) yield correction other than .00. Using the final loadings of .90, .60, and .40 to recompute the residual table yields

	1	2	3
1	——	.000	.000
2	.000	——	.000
3	.000	.000	——

indicating that we have achieved the exact solution sought for.

Note: Actually the copying of the residual table each time would greatly increase the labor and is not necessary. If room is allowed in the original table for several new entries in each residual cell, all of the process can be carried out in one half of the original table. The present problem would then look as follows [The reader can follow by performing the same steps as above in the given order]:

	1	2	3	Loadings
1	——	.022 —.047 (2) .038 (3) .008 (4) —.001 (5)	—.015 .060 (1) .020 (2) .000 (4)	.75 (+.10) (2) .85 (+.05) (4) .90
2		——	—.105 —.036 (1) .004 (3) .000 (5)	.69 (—.10) (3) .59 (+.01) (5) .60
3			——	.50 (—.10) (1) .40

Sums of	.007	—.083	—.120* (1)
	.082* (2)	—.016	.024
Residuals	—.027	—.083* (3)	.016
	.058* (4)	.042	.016
	.008	.012* (5)	.004
	—.001	.001	—.000

A Two-Factor, Four-Variable Problem

To further illustrate the method and to demonstrate that it will work for problems containing more than one factor, a second analysis of four tests containing two common factors is presented.

The Thurstone centroid method, using the highest intercorrelation in row or column as the estimate of the communality, yielded:

.48	.48	.32	.00		
.48	*.48*	.48	.36		
.32	.48	*.48*	.24		
.00	.36	.24	*.36*		
1.28	1.80	1.52	.96	5.56	2.36
.54	.76	.64	.41		2.35

		(—)	(—)		
.22	.07	.03	.22		
.07	*.07*	.01	—.05		
.03	.01	*.03*	—.02		
.22	—.05	—.02	*.22*		
.54	.10	.05	.37	1.06	1.03
.52	.10	.05	.36		(1.03)

		(—)	(—)	
—.05	.02	.00	.03	[.00]
.02	*.06*	.00	—.09	[—.01]
.00	.00	*.03*	—.04	[—.01]
.03	—.09	—.04	*.10*	[.00]

The centroids were rotated to simple orthogonal structure, and the residuals recomputed (to three decimal places this time) to serve as a more accurate basis for correction. The results of this were

Centroid Loadings		Rotated Loadings			Residuals after Rotation			
I	*II*	*A*	*B*		1	2	3	4
.54	.52	.07	.75	1	——	.017	—.001	—.031
.76	.10	.51	.57	2		——	—.002	.085
.64	—.05	.52	.38	3			——	—.042
.41	—.36	.55	—.01	4				——

The application of the iterative correction method is shown in Table 1. The successive sequences of action are all numbered. The four steps in each sequence are: (a) noting of test with most unbalanced sum of residuals, (b) determining which factor pattern is most like the pattern of residuals and how much correction to apply, (c) actually making the correction to the factor loading and in the residual table, and (d) computing the new sum of residuals for each column. The reader is left to follow these steps for himself as an exercise in the method.

The original rotated loadings, the corrected loadings, and the re-rotated loadings are:

TABLE 1
Application of the Iterative Correction Procedure to a Two-Factor, Four-Variable Problem

	1	2	3	4	Factor Loadings A	Factor Loadings B
1		.017 .010 (1) .007 (3) .033 (4) —.001 (5) .000 (7) .006 (8) .000 (9)	—.001 .006 (2) .023 (4) .000 (5) .004 (8) .000 (9)	—.031 —.009 (4) —.008 (5) —.009 (6) —.003 (8)	.07 (—.04) (4) .03 (—.01) (8) .02	.75 (+.06) (5) .81 (+.01) (9) .82
2			—.002 —.054 (1) .007 (2) —.010 (3) —.002 (7)	.085 .030 (1) .008 (3) —.012 (6) .000 (7)	.51 (+.10) (1) .61 (+.04) (3) .65 (—.02) (7) .63	.57
3				—.042 .013 (2) .000 (6)	.52 (—.10) (2) .42	.38
4					.55 (+.03) (6) .58	.01
Sum of Residuals	(0) —.015 (1) —.022 (2) —.015 (3) —.018* (4) (4) .045* (5) (5) —.009 (6) —.010 (7) —.009* (8) (8) .007* (9) (9) —.003	.100* (1) —.014 .047* (3) .005 .031 —.003 —.023* (7) —.002 .004 —.002	—.047 —.097* (2) .026 .009 .026 .003 —.010 —.002 .002 —.002	.012 —.043 .012 —.010 .012 .013* (6) —.021 —.009 —.003 —.003		

Test	Original Rotated Loadings		Corrected Loadings		Re-rotated Loadings	
	A	B	A^1	B^1	A^{11}	B^{11}
1.	.07	.75	.02	.82	—.02	.82
2.	.51	.57	.63	.57	.60	.60
3.	.52	.38	.42	.38	.40	.40
4.	.55	—.01	.58	—.01	.58	.02

The reader might note that original centroid loadings for tests 2 and 3 on Factor A were in considerable error, with variable 3 made to appear to have equal projection along with variables 2 and 4.

The re-rotated loadings yield residuals of

	1	2	3	4
1	—	.000	.000	—.005
2		—	.000	.000
3			—	.000
4				—

TABLE 2
Further Application of the Iterative Correction Procedure to Remove Small Remaining Residuals

	1	2	3	4	A^1	B^1
1		.000	.000	-.005	—.02(+.02)(3)	.82(—.02)(4)
		-.012(3)	-.008(3)	.011(1)	.00	.80
		.000(4)	.000(4)	-.001(3)		
2			.000	.000	.60	.60
				.012(1)		
				.000(2)		
3				.000	.40	.40
				.008(1)		
				.000(2)		
4					.58(+.02)2	.02(—.02)(1)
					.60	.00
(0)	-.005	.000	.000	-.005*(1)		
(1)	.011	.012	.008	.031*(2)		
(2)	.011*(3)	.000	.000	.011		
(3)	-.021*(4)	-.012	-.008	-.001		
(4)	-.001	.000	.000	-.001		

and even this one small residual can be eliminated if one wishes to pursue the matter through the four added steps in Table 2. Actually the procedure in Table 2 is of no consequence, involves seeing several

steps ahead (as in chess), and is presented only to show that the present method can and will give, in a theoretical problem, absolutely correct results if pursued far enough. The final factor loadings of

A^{11}	B^{11}
.00	.80
.60	.60
.40	.40
.60	.00

are the theoretical values from which the table was set up and of course give residuals of zero.

It should be noted that the two examples presented were both theoretical problems, which accounts for the fact that all residuals finally became zero. In actual problems the residuals become small, but do not disappear.

Practical Application

The writer, and technicians at the Personnel Research Section, AGO, Department of the Army and at the Medical Research Department, U. S. Submarine Base, have applied the method successfully to dozens of factorial solutions, one problem containing 39 variables and 13 factors. It has consistently lead to lower residuals, better balanced residuals, fewer *queer* factor loadings, and more exact duplication of factor structure in cross-validation studies.

Another use, suggested by Richard H. Gaylord, Personnel Research Section, AGO, and successfully tested, at least on one problem, by the author, is its use in cross-validation factor studies. In the case to which it was applied, four sets of data consisting of the same 11 variables had been collected on four different populations, each from a widely different geographical area. One set of data was factor-analyzed and the factors rotated and then improved by the present technique. These loadings were in turn *assumed* to apply to the other three sets of intercorrelations and residual tables were computed for them. In each set the loadings were then corrected by the iterative method described above until the residuals became small and balanced. It is estimated that about three-fourths of the usual time was saved on these three analyses.

Another use to which it can be put is adding one or two extra variables to an already completed analysis when to apply Dwyer's extension would take much longer. Consider the addition of a fifth variable to the four-variable problem solved above. Let us suppose that its correlations with the other variables are $r_{15} = .16$, $r_{25} = 36$, $r_{35} = .24$, and $r_{45} = .24$. If we consider that this variable is in the matrix,

its loadings are now zero for both factors and its correlations with the tests are its residuals. We then have

	1	2	3	4	5	A^{11}	B^{11}
1	—	.00	.00	.00	.16 .00(2)	.00	.80
2		—	.00	.00	.36 .12(1) .00(2)	.60	.60
3			—	.00	.24 .08(1) .00(2)	.40	.40
4				—	.24 .00(1)	.60	.00
5					—	.00(+.40) (1) .40	.00(+.20) (2) .20

Thus two simple estimations, 6 multiplications, and 6 subtractions replace the squaring, getting of cross-products, the Doolittle forward solution and two back solutions, as well as yielding the residuals as an instantaneous check on the accuracy of the fit.

While the presentation has been made in terms of orthogonal factors—to which the writer is addicted—and while the method can be applied only to factors in the orthogonal form, the method does not prevent applying the principle of simple structure and the securing of oblique factors *after* the process is completed. Since most methods yield an orthogonal set of factors initially, and since group or cluster procedures usually rotate to orthogonality as an intermediate step, at least, the present method of correction should be useful in all cases without involving much added work.

In summary, the present iterative method, while preserving orthogonality by always postulating $R=FF'$, more nearly satisfies the criterion of zero residuals (other than chance) by testing the chance hypothesis a row (or column) at a time instead of applying it to the whole table at once. It is applicable to any method which assumes communalities in the diagonals and can be viewed as a correction to the original estimates of such communalities.

PSYCHOMETRIKA—VOL. 14, NO. 3
SEPTEMBER, 1949

A SIMPLIFIED PUNCH CARD METHOD OF DETERMINING SUMS OF SQUARES AND SUMS OF PRODUCTS

GEORGE F. CASTORE
DEPARTMENT OF PSYCHOLOGY, COLGATE UNIVERSITY

AND

WILLIAM S. DYE, III
TABULATING DEPARTMENT, THE PENNSYLVANIA STATE COLLEGE

A simplified method of obtaining sums of squares and sums of cross products by the use of punch card equipment is described. Application of the method has revealed several advantages, which are noted.

Of the many procedures in use to obtain the sums of squares and sums of products which are needed in various equations in statistical formulas, progressive digiting through the use of punch cards is one of the faster and more accurate ways. The method presented here is considered as a special adaptation of the traditional punch card method which increases its efficiency and simplifies the operation.

The advantages of this proposed method may be listed briefly before the method is described so that the reader may be able to picture them as a function of the procedure. The advantages are:

1. There is only one card sort for each variable and one change of controls during the digiting for each variable.
2. Units, tens, and hundreds controls are a single tabulation process, which saves time by decreasing control breaks for printing and punching.
3. No rewiring is necessary for increasing the number of variables from 10 to 25.
4. A permanent board may be wired, thereby saving time in the wiring for each new digiting job.
5. Different groups may be digited at the same time by gang-punching different numbers in one column for each group and major-controlling on that column.
6. Summarizing gives the complete sums of squares and sums of products. No further addition by hand or calculator is required to obtain these sums. An additional advantage is that errors due to sorting, wiring, or the machine may thus be immediately detected and corrected.

Necessary Equipment

In order to use this method it is necessary to have all data on punch cards and to have available an eighty-counter tabulator equipped with eighty counters for progressive totalling, a reproducer or summary punch, and a sorting machine.*

Conditions Imposed on Discussion, Tables, and Figures

In order to simplify the explanations, certain conditions have been imposed upon the material as presented in this discussion. None of the limitations are an actual restriction of the method, which is flexible enough to meet the requirements of an increase in the number of variables or in the size of scores. The imposed conditions are that (1) no summation of all the raw scores for any given variable exceeds 90,000; (2) the largest single raw score is a three-place figure, e.g., 365 or 942; (3) the number of variables is 10.

Arrangement of Data on Punch Cards

Execution of this method depends upon a specific arrangement of raw scores punched on cards. The original punched-card data must be reproduced completely on three separate decks. The three decks will be used in the digiting process.

Variables are placed in fields of five columns each. The field for the first variable would be from column 6 to 10, inclusive, and so on for each of ten variables. The fields are identical for each of the three decks.

Assume that an individual's score for variable no. 1 is 123 and his score for variable no. 2 is 136. On deck number 1 (1 is gang-punched in column 1 of this deck, which is the units deck) 123 is puched in columns 8, 9, and 10, respectively, and 136 is in columns 13, 14, and 15, respectively. In deck number 2 (2 is gang-punched in column 1 of the tens deck), each score is set off one place to the left so that 123 is punched in columns 7, 8, and 9 and 136 is columns 12, 13, and 14. In deck number 3 (3 is gang-punched in column 1), each score is set off two places from the right column so that 123 is now punched in columns 6, 7, and 8 and 136 appears in columns 11, 12, and 13. The arrangement would appear as shown in Table 1. This plan makes possible the totaling of all squares and cross-products with a single machine process; i.e., it makes unnecessary the controlling on the units, tens, and hundreds position separately.

*The machines which were used here were a type 405 Alphabetic Accountting Machine, a type 513 Reproducing-Summary Punch, and a type 080 Sorting Machine.

TABLE 1
Arrangement of Two Variables on Each of Three Decks in Selected Fields and Controls Which Improved Digiting Process

Deck	Field Variable I					Columns Variable II					Control Var. I	Columns Var. II
Number	6	7	8	9	10	11	12	13	14	15	56	57
1 (units)			1	2	3			1	3	6	3	6
2 (tens)		1	2	3			1	3	6		2	3
3 (hundreds)	1	2	3			1	3	6			1	1

Columns 56 to 65, inclusive, are used for the variable controls. Column 56 on each card of each deck contains the controls for variable 1; column 57 contains controls for variable 2; column 58 controls for variable 3, etc. Deck number 1 contains all unit controls. Deck number 2 contains all tens controls, deck number 3 contains all hundreds controls. Thus, using the former example of a variable 1 score as 123 and a variable 2 score as 136, deck number 1 would have a 3 in column 56 and a 6 in column 57. Deck number 2 would have a 2 in column 56 and a 3 in column 57. Deck number 3 would have a 1 in column 56 and a 1 in column 57. (See Tables 1 and 2). The numbers are punched into columns 56 to 65, inclusive, by use of a split wire from the units position of the original data card when the three decks are reproduced. The split wire is shifted to the tens position for each variable when deck number 2 is reproduced and to the hundreds position for deck 3.

TABLE 2
Fields* of Placement on Punch Cards for Ten Variables, and Sorting and Digiting Controls to be Used with Imroved Digiting Process

Var. No.	Field Columns (inclusive)	Columns of Field for Punching: Units Card	Tens Card	Hundreds Card	Control Column
1	6 to 10	8, 9, 10	7, 8, 9	6, 7, 8	56
2	11 to 15	13, 14, 15	12, 13, 14	11, 12, 13	57
3	16 to 20	18, 19, 20	17, 18, 19	16, 17, 18	58
4	21 to 25	23, 24, 25	22, 23, 24	21, 22, 23	59
5	26 to 30	28, 29, 30	27, 28, 29	26, 27, 28	60
6	31 to 35	33, 34, 35	32, 33, 34	31, 32, 33	61
7	36 to 40	38, 39, 40	37, 38, 39	36, 37, 38	62
8	41 to 45	43, 44, 45	42, 43, 44	41, 42, 43	63
9	46 to 50	48, 49, 50	47, 48, 49	46, 47, 48	64
10	51 to 55	53, 54, 55	52, 53, 54	51, 52, 53	65

*Columns 1 to 5 usually contain identification information. Deck number is usually in column 1.

In addition to the work deck, ten cards should be key-punched, all

having "X" in column 80. One card should have "zero's" in columns 56 to 65, another should have "1's" in columns 56 to 65, etc., up to "9's" in columns 56 to 65. These are to be sorted in with the digiting decks. Their purpose is to leave no blank pockets in the sorter while sorting is being done.

The cards are now in a form such that we need only sort once on each variable column, cut the progressive total summary cards, and retotal these summaries to obtain the required sums.

The Digiting Process

The 405 Alphabetic Accounting Machine Board should be wired as follows:

1. All counters progressive, clear progressive on major total cycle.
2. Seventy counters coupled together and made to add on No 80X.
3. Wire the ten variables from add brushes to counters so as to have two extra counters to the left of each variable. Total print in any way, but clear all total exit wires for each variable.
4. Wire unequal impulse to major shunt and also to digit pick-up of a class selector.
5. Wire column 56 (upper and lower brushes) to automatic control. These wires are to be moved for each successive variable after the first run. Wire unequal impulse to minor shunt.
6. Clear counter on minor class total.
7. To get line control for summary punching, pick up an X distributor on *"Digit"* with first card minor. Run a *hot* 9 through controlled points and to the No X side of the class selector mentioned in step 4. The controlled side of this selector is connected to lower brush column 56, which is split-wired into this position and automatic control. The common side goes to a tens position of a two-position counter.
8. Variable number may be wired to a two-position counter from numbers in the digit selector, the wires being changed for each variable. This may be controlled also by first card minor.

The 513 board should be wired to gang-punch the variable number in columns 2 and 3 and to summary-punch the progressive totals for the ten variables in columns 6 to 75.

After the totals for all variables have been totaled for each deck for accuracy, sort all three decks along with the ten prepared cards on column 56, the position of the sorting variable number 1. Remove cards from the sorter in descending order from 9 to 0. The prepared

extra ten cards provide assurance that there will be a card in every pocket of the sorter and that there will be a control break and a progressive total and summary card for each number during tabulation. It is not necessary to use any of the ten extra cards which are higher than the highest detail card of the variable which is being sorted upon.

To illustrate what is taking place step by step, assume that there are nine detail cards reproduced from three original cards containing the following data for variables 1 and 2. (Remember there are three cards for each example, each with variable punches set off one place.)

Original Data Card No.	Score for Variable 1	Score for Variable 2
1	123	136
2	234	345
3	345	234

Table 3 shows the order of the above nine cards and arrangement of data from the first card to the last following the first sort in column 56. It may be noticed that units, tens, and hundreds controls operate together.

TABLE 3
The Order of Nine Cards and Corresponding Raw Scores Following the Sort on Control Column 56 for the Two Variables Given in the Example

	Columns for Raw Scores										Sort-Control Columns	
Deck	Variable I					Variable II					Var. I	Var. II
Number	6	7	8	9	10	11	12	13	14	15	56	57
1			3	4	5			2	3	4	5	4
1			2	3	4			3	4	5	4	5
2		3	4	5			2	3	4		4	3
1			1	2	3			1	3	6	3	6
2		2	3	4			3	4	5		3	4
3	3	4	5			2	3	4			3	2
2		1	2	3			1	3	6		2	3
3	2	3	4			3	4	5			2	3
3	1	2	3			1	3	6			1	1

Table 4 shows the progressive totals that will be printed and punched on summary cards for the nine cards. Following this tabulation, sort all cards on column 57. Remove the cards from the sorter in descending order. Move the control wires from column 56 to 57. Move the units wire of the variable control to *digit* 2. The cards are now in the order shown in Table 5. Notice that the order of the cards is determined by the sorting variable at the extreme right of Tables 3

TABLE 4

Progressive Totals as They Will Print and Summary-Punch for the Nine Cards in the Example When Sorted on Column 56 and Tabulated

Variable Number	Group Indication	Progressive Totals Variable I	Variable II
1	5	345	234
1	4	4029	2919
1	3	40992	29905
1	2	65622	65765
1	1	77922	79365
1	0	77922	79365

TABLE 5

The Order of Nine Cards and Corresponding Raw Scores Following the Sort on Control Column 57 for the Two Variables Given in the Example

	Columns for Raw Scores										Sort-Control Columns	
Deck	Variable I					Variable II					Var. I	Var. II
Number	6	7	8	9	10	11	12	13	14	15	56	57
1			1	2	3			1	3	6	3	6
1			2	3	4			3	4	5	4	5
1			3	4	5			2	3	4	5	4
2		2	3	4			3	4	5		3	4
2		3	4	5			2	3	4		4	3
2		1	2	3			1	3	6		2	3
3	2	3	4			3	4	5			2	3
3	3	4	5			2	3	4			3	2
3	1	2	3			1	3	6			1	1

TABLE 6

Progressive Totals as They Will Print and Summary-Punch for the Nine Cards in the Example When Sorted on Column 57 and Tabulated

Variable Number	Group Indication	Progressive Totals Variable I	Variable II
2	6	123	136
2	5	357	481
2	4	3042	4165
2	3	31122	42365
2	2	65622	65765
2	1	77922	79365
2	0	77922	79365

TABLE 7

Final Totals When Summary Cards as Shown In Tables 4 and 6 Are Totaled

Variable Number	Final Variable I	Totals Variable II
01	188910	178188
02	178188	192277

and 5 and that units, tens, and hundreds controls are totaling at the same time. Table 6 shows the progressive totals which are printed and summary-punched for the nine cards sorted and controlled on column 57.

The summary cards were punched for this two-variable example as shown in Tables 4 and 6. *By excluding all summary cards having zero for group indication* and adding the remaining cards while minor controlling on columns 2 and 3 (variable columns), the final sums should be obtained as shown in Table 7. Notice the cross checks which indicate that the tabulation is accurate, 178188.

Checks for Accuracy

The sums of squares and cross products are printed for all ten variables in the final tabulation. There is a cross check on every cross product (note figures 178188 in Table 7) which should serve as an indication not only that no card is missing from a variable group but also that the machine is functioning properly for each vertical column. If all the cross products cross-check the summations of the squares are assumed to be accurate.

In studies in which this method has been used the actual sorting and tabulation for ten variables and an N of two to three hundred takes approximately two hours.

Flexibility

If the number of variables is increased to 20, it is necessary only to make another set of three decks as the first three were made. *The caution here is that the controls for the second set must be on the first set and second set in columns* 66 *to* 75 *and the controls on the first set should be duplicated on the second set in columns* 56 *to* 65.

If the summation of scores for any variable exceeds 90,000, either of the two following adjustments may be made: (1) allow two fields for one variable or (2) allow 8 progressive-totaling columns for the first five variables and seven for the last five variables.

If scores exceed three-place figures, fields may be doubled or enlarged. Four-place figures may be reduced to three-place figures by subtracting a constant if the range does not exceed 999.

If two or more groups are being tabulated at once each group must be sorted separately although tabulation may be continuous.

PSYCHOMETRIKA—VOL. 14, NO. 2
JUNE, 1949

CONSTITUTION OF THE PSYCHOMETRIC SOCIETY

ARTICLE I

Object

The primary purpose of the Psychometric Society is to promote the development of psychology as a quantitative rational science. This concept of quantification involves the formulation of hypotheses in mathematical form, their development into a consistent quantitative psychological theory, and quantitative tests of the agreement between theory and experimental data.

ARTICLE II

Membership

1. Members of the Society shall be persons who are interested in the development of psychology as a quantitative rational science and who, from their training and experience, give evidence of their ability to contribute either directly or indirectly to the objectives of the Society as set forth in Article I.

2. Members shall be entitled to recieve such printed matter and to participate in such scientific meetings as the Society may direct.

3. A membership may be terminated at any time by a majority vote of the Members at any Annual Meeting upon recommendation of the Council of Directors after investigation.

4. Members shall be elected by a majority vote of the members upon recommendation by the Membership Committee and nomination by the Council of Directors.

5. The Council of Directors shall have power to defer action upon such proposals for membership as it deems necessary, provided, however, that by the second Annual Meeting after the original request for membership it must decide either to present or not to present the nominee's name to the Society. A proposal for membership cannot be renewed until one year has elapsed after the Council's action upon it.

6. A member shall be allowed to vote only if he has paid all his annual dues from the time of his election to membership.

ARTICLE III

Meetings

1. The Annual Meeting of the Members of the Society shall be held at a time and place determined by the Society.

2. Other meetings may be held upon call of the Executive Committee.

3. A quorum shall consist of twenty Members in good standing.

ARTICLE IV

Council of Directors

1. The Council of Directors shall consist of the President, the Secretary, the Treasurer, and six Directors chosen from the membership. Each of the Directors shall serve for a term of three consecutive years. Two of the Directors shall be elected each year, as provided in Section 2 of this article. The terms of all Directors shall begin October first and shall expire September thirtieth.

2. The council of Directors shall nominate from the membership, by majority vote, two or more candidates to replace the retiring Directors, and shall publish the names of the candidates in Psychometrika at least two months before the Annual Meeting of the Society. Nominations for each candidate, other than those nominated by the Council of Directors, may be made by a signed petition from ten or more Members of the Society. Such signed petitions must be received by the Secretary not less than thirty days prior to the day of the Annual Meeting.

3. If only two candidates are nominated they shall be elected by viva voce vote at the next Annual Meeting. If more than two candidates are nominated the election shall be by preferential vote at the Annual Meeting.

4. No person may serve two consecutive three-year terms as one of the six Directors.

5. If any of the Directors be elected an Officer of the Society his term as Director automatically expires. The Council of Directors shall appoint a Member of the Society to fill the unexpired portion of any Directorship terminated before the expiration of the three-year term.

6. The Council of Directors shall exercise general supervision of the affairs of the Society, shall nominate new Members upon recommendation of the Membership Committee, and shall make recommendations concerning the conduct of the Society which shall be brought before the Members of the Society at any duly constituted meeting. The Council of Directors shall have the power to make such contracts and to provide for the delivery of such instruments as shall be necessary for the carrying out of all purposes, functions and other business of the Society as shall be authorized by vote of the Members of the Society at any duly constituted meeting, or as may be elsewhere provided by this Constitution.

7. The President of the Society shall be Chairman ex officio of the Council of Directors, and the Secretary of the Society shall be Secretary ex officio of the Council of Directors.

Article V

Officers

1. The officers of the Society shall be: a President, a Secretary, and a Treasurer. These officers shall constitute the Executive Committee. The terms of all officers shall begin October first and shall expire September thirtieth.

2. All officers of the Society must be Members of the Society and also Members or Associates of the American Psychological Association.

3. The President shall be elected for the term of one year, and shall not hold office for two successive terms. The Secretary and Treasurer shall be elected for terms of three years each.

4. The President shall be elected annually by the Society, a nominating and an election ballot being successively cast under the supervision of the Election Committee as provided in Article VI of this Constitution. Election shall be by means of a preferential voting system.

5. The Secretary and the Treasurer shall be elected by a majority vote of the Members present at an Annual Meeting, upon nomination by the Council of Directors.

6. It shall be the duty of the President to preside at all meetings of the Society, to act ex officio as Chairman of the Council of Directors, to countersign all contracts and other instruments of the Society except checks, to exercise gen-

eral supervision over the affairs of the Society and to perform all such other duties incident to his office or required of him by vote of the Members of the Council of Directors at any duly constituted meeting. The first President of the Society shall be the first Chairman of the Editorial Council and shall make such appointments as are necessary for carrying out the provisions of this Constitution.

7. It shall be the duty of the Secretary to keep the records of all meetings of the Society and of the Council of Directors; to file and hold subject to call, and to arrange for the publication of such records, reports, and proceedings as are authorized by this Constitution, and also by vote of the Members or of the Council of Directors at any duly constituted meeting; to bring to the attention of the Council of Directors and the Society such matters as he deems necessary; to conduct the official correspondence of the Society and the Council of Directors; to have custody of such bonds as may be required to be filed by the Treasurer and such other employees as shall be required by the Society to file a bond, holding these bonds subject to the order and direction of the Society; to issue notices of meetings; to assume in the case of the death or incapacity of the President the duties of the President until such time as a successor is appointed by the Council of Directors or elected by the Members; to sign such checks or other drafts upon the funds of the Society as may be necessary in case of the death or incapacity of the Treasurer, and the Secretary is hereby authorized to sign such checks or drafts in such contingency; to execute or deliver any documents which he shall be directed to execute or deliver on behalf of the Society by the Constitution, vote of the Members of the Society, or the Council of Directors; and in general to perform all such other duties as are incident to his office or as properly may be required of him by vote of the Members or the Council of Directors at any duly constituted meeting. In the absence of any specific provision of this Constitution to the contrary, the Secretary shall have power and authority to represent the Society in the voting or other management of any stock held by the Society in any corporation or company; and in the event that the performance of such acts by the Secretary becomes impossible or inadvisable, by virtue of law or otherwise, the Secretary shall have the power to appoint any Member of the Society to act as duly authorized agent of the Society for the performance of said acts.

8. It shall be the duty of the Treasurer to have custody of all funds, stocks, securities and to deposit the same in the name of the Society in such bank or banks as the Society or Council of Directors may direct; to have custody of all other property of the Society not otherwise expressly provided for by this Constitution and to hold same subject to the order and direction of the Society; to collect dues and other debts due the Society by all persons and organizations whatsoever; and to execute or deliver any documents which he shall be directed to execute or deliver on behalf of the Society by the Constitution, vote of the Members or the Council of Directors. He shall have the authority to sign checks and drafts on behalf of the Society for the disbursement of funds for the duly authorized purposes of the Society as provided by the Constitution, vote of the Members of the Society, or Council of Directors. He shall be bonded for an amount fixed by the Council of Directors, the bond to be filed with the Secretary. He shall, at all reasonable times, exhibit his books and accounts to any Member of the Society. He shall keep a full and complete record of all money received and all money paid out, and shall perform such other duties as may be reasonably required of him by vote of the Members of the Society at a duly constituted meeting, or by the Council of Directors.

9. In case of the death, disability, or resignation of any of the officers, the Council of Directors shall appoint a successor to serve until the next Annual Meeting of the Society. Vacancies existing at the time of an Annual Meeting shall be filled by vote of the Members at the meeting as provided by Sections 4 and 5 of this article.

ARTICLE VI

Elections

1. The Secretary shall issue a call for nominating ballots for the nomination of President at least five months before the Annual Meeting. The ballots shall be counted by the Election Committee at least four months before the Annual Meeting. In case any nominee receives a majority of first choices on the nominating ballot he shall be declared elected. Otherwise the Election Committee shall send to all Members a ballot containing the names of not less than two nor more than five nominees receiving a large number of votes. This ballot shall contain two blank spaces in which names may be written. The votes shall be counted at least one month before the Annual Meeting, and the results announced at the Annual Meeting.

ARTICLE VII

Committees

1. The Committees of the Society shall consist of such standing committees as may be provided in this Constitution and such special Committees as may be established by vote of the Members or the Council of Directors, or as may be appointed by the President.

2. It shall be the duty of the Executive Committee to make the arrangements necessary for the Annual Meeting and for any other meetings which the Society may authorize; to bring to the attention of the Council of Directors any matters, not specifically provided for by this Constitution, requiring action by the Council; and to perform such other functions incident to the activity of the Society as may be reasonably required of it.

3. The Program Committee shall consist of three members, each of whom shall serve for three years. Each year, before two months shall have elapsed since the last Annual Meeting, the President shall appoint one member to the Committee to take the place of the retiring member. No member of the Committee may serve two successive terms. The members of the Committee shall be appointed by the President with the approval of the Council of Directors. The chairman of the Committee shall be the member who has served the longest.

4. The Election Committee shall consist of three members, each of whom shall serve for three years. Each year, before two months shall have elapsed since the last Annual Meeting, the President shall appoint one member to the Committee to take the place of the retiring member. The chairman of the Committee shall be the member who has served the longest. It shall be the duty of the Election Committee to conduct and supervise all elections of the Society.

5. The Membership Committee shall consist of three members, each of whom shall serve for three years. Each year, before two months shall have elapsed since the last Annual Meeting, the President shall appoint one member to the Committee to take the place of the retiring member. The chairman of the Committee shall be the member who has served the longest. It shall be the duty of the Membership Committee to invite persons qualified in the sense of Article II,

Section 1, to apply for membership, to provide the necessary forms, and to recommend qualified persons for membership to the Council of Directors.

Article VIII

Publications

1. The official publication of the Society shall be Psychometrika.

2. The Society shall pay to the Psychometric Corporation ninety per cent of the annual dues collected from the Members, in return for which it shall accept for each Member a subscription to Psychometrika for that year. Such monies shall be payable to the Psychometric Corporation within three months after receipt by the Society.

3. The Society shall provide an Editorial Council for the publication of Psychometrika.

Article IX

Editorial Council

1. The Editorial Council shall consist of a Chairman, two Editors, and a Managing Editor. Each year the Editorial Council shall appoint for a term of two years such members as it may desire to an Editorial Board which shall assist the Editorial Council in the publication of Psychometrika.

2. The Chairman of the Editorial Council shall be appointed for a term of five years. The Editors and the Managing Editor shall each be appointed for a term of three years in a manner such that the term of one of these three expires each year.

3. New members of the Editorial Council shall be appointed by a majority vote of the Council of Directors with the approval of the Psychometric Corporation.

Article X

Annual Dues

1. The annual dues for Members shall be five dollars a year, payable January first of each year. Non-payment of dues for two consecutive years shall be considered equivalent to resignation from the Society.

2. Any Member, upon payment of the dues prescribed by this article, shall receive Psychometrika without further charge, throughout that membership year to which the dues shall be applicable.

3. New Members shall pay dues for the entire calendar year in which they are elected, and shall receive copies of Psychometrika for the entire year, if such copies are available.

Article XI

Scientific Programs

1. The scientific programs of the Society shall be conducted and supervised by the Program Committee. This Committee shall have full power in the selection and rejection of papers, for all scientific programs, provided that:

2. When meetings are held jointly with other societies the program as a whole shall be subject to the approval of such societies.

Article XII

Amendments

1. This Constitution may be amended by a vote of two-thirds of the Mem-

bers present at any Annual Meeting or by a two-thirds vote of all Members responding by vote to a mailed ballot, provided that:

2. The proposed amendment shall have been previously approved by a three-fourths vote of the entire membership of the Council of Directors and the Editorial Council as a whole.

AMENDMENT I

Student members of the Society shall be persons who are studying and preparing for professional development in line with the objectives of the Society. The student status of each applicant for this type of membership shall be certified annually by the registrar or by a faculty member of his college or university. Student members shall have all rights and privileges of the Society except those of voting and holding office. Under no circumstances shall the privilege of student membership be granted for a period exceeding three years. The annual dues for student members shall be $3.00 a year, payable January 1 of each year.

PSYCHOMETRIKA—VOL. 14, NO. 3
SEPTEMBER, 1949

BOOK REVIEWS

CHURCHMAN, C. WEST *Theory of Experimental Inference.* New York: Macmillan Company, 1948. Pp. 292.

In this revision and expansion of an earlier experimental edition, the author points out the important presuppositions of inquiry, and he attempts to relate the problems of scientific methodology to a fundamental philosophy of science.

There are sixteen chapters in the book, and a listing of their titles may be pardoned as a means of indicating the general coverage of the work: On the Nature of Statistical Tests; General Methodology of Inference; Problems of Method; Dialectic of Modern Philosophy; Relationalism; Naive Empiricism; Statistical Empiricism; Criticism; Relativism; Experimentalism I—The Answering of Questions; Experimentalism II — On Meaning and Method; Experimentalism III — Nonmechanical Concepts; Applications of Experimentalism; On Science, Personality, and Social Conflict; On Chance, Loss, and Risk; On Quality Control — An Ideal.

It will be seen by the scope of this small book that Churchman has attempted to unravel a snarl of many threads and then weave them into a nice tapestry, one which even has social significance. In some cases, however, the strands chosen for exhibition are far too short for the purpose. The nature of statistical tests is "covered" in 13 pages where many of the major concepts of statistical inference are mentioned, but of course not explained. If the reader is a statistician he does not need the first three chapters in their present form, and if he is not a statistician these chapters will be merely confusing. The reviewer informally tested this hypothesis to his satisfaction on two available non-random engineers.

In the remaining chapters, however, Churchman presents a stimulating and provocative account of the philosophical backgrounds of modern concepts of scientific methodology. These sections are judged by this reviewer as very useful analyses, to be read by all serious students of science. This work is not easily read, and the important implications will not be arrived at by scanning. Admiration and real results do, however, come with a little labor. The main thesis presented is one on which most experimental researchers in psychology will agree, but they will now be able to acknowledge *why* they agree.

In the last chapter but one, there is presented an important analysis of scientific and non-scientific problem-solving in terms of social phenomena, motivation, and ethical considerations. The general field of quality control is considered in the last chapter in terms of means-end relations, that is, scientific method and social effects.

To recapitulate, this reviewer feels that Churchman has produced an extremely important analysis and synthesis of many of the problems of scientific methodology. Researchers in the human sciences especially will profit from a deep and careful reading of this book.

Department of the Army T. G. ANDREWS